International Society for Photogrammetry and Remote Sensing
Société Internationale de Photogrammétrie et de Télédétection
Internationale Gesellschaft für Photogrammetrie und Fernerkundung

THE INTERNATIONAL ARCHIVES OF THE PHOTOGRAMMETRY, REMOTE SENSING AND SPATIAL INFORMATION SCIENCES
ARCHIVES INTERNATIONALES DES SCIENCES DE LA PHOTOGRAMMÉTRIE, DE LA TÉLÉDÉTECTION ET DE L'INFORMATION SPATIALE
INTERNATIONALES ARCHIV FÜR PHOTOGRAMMETRIE, FERNERKUNDUNG UND RAUMBEZOGENE INFORMATIONSWISSENSCHAFTEN

VOLUME / VOLUME / BAND XXXVIII

PART / TOME / TEIL 7A

ISPRS Technical Commission VII Symposium

100 Years ISPRS
Advancing Remote Sensing Science

Vienna, Austria
July 5 – 7, 2010

Papers accepted on the basis of peer-reviewed full manuscripts

Editors
W. Wagner, B. Székely

Organised by
ISPRS Technical Commission VII
Thematic Processing, Modeling and Analysis of Remotely Sensed Data

Supported by
Leica Geosystems, Vexcel Imaging, Stadt Wien, ESRI, RIEGL LMS, TU Wien

ISSN 1682-1777

Credits

Technical Editing: Alexandra von Beringe, Peter Dorninger, Sebastian Flöry, Josef Jansa, Clemens Nothegger, Norbert Pfeifer, Andreas Roncat

using LaTeX's confproc class (by V. Verfaille)

Cover Illustration: Gregor Franzen

Printed by: Buchdruckerei Ernst Becvar Ges.mbH, Vienna, Austria

Published by
Institute of Photogrammetry and Remote Sensing, Vienna University of Technology
on behalf of
ISPRS Technical Commission VII *Thematic Processing, Modeling and Analysis of Remotely Sensed Data*
Cooperating ISPRS Working Groups
- WG VII/1 *Physical Modelling and Signatures in Remote Sensing*
- WG VII/2 *SAR Interferometry*
- WG VII/3 *Information Extraction from Hyperspectral Data*
- WG VII/4 *Methods for Land Cover Classification*
- WG VII/5 *Methods for Change Detection and Process Modelling*
- WG VII/6 *Remote Sensing Data Fusion*
- WG VII/7 *Theory and Experiments in Radar and Lidar*
- ICWG III/VII *Pattern Recognition in Remote Sensing*

ISPRS Headquarters 2008-2012
c/o CHEN JUN, ISPRS Secretary General
National Geomatics Centre of China
No. 1 Baishengcun, Zizhuyuan
Beijing 100048, PR CHINA
Tel: +86 10 6842 4072
Fax: +86 10 6842 4101
Email: chenjun@nsdi.gov.cn; chenjun_isprs@263.net
ISPRS WEB Homepage: http://www.isprs.org

Available from
GITC bv
P.O.Box 112
8530 AC Lemmer
The Netherlands
Tel: +31 (0) 514 56 18 54
Fax: +31 (0) 514 56 38 98
E-mail: mailbox@gitc.nl
Website: www.gitc.nl

Table of Contents

Preface

These are the proceedings of the ISPRS Technical Commission VII Symposium that was held on July 5–7, 2010, at the Vienna University of Technology, Austria. The proceedings consist of two parts: Part A collects all papers that were accepted on the basis of peer-reviewed full manuscripts; Part B contains papers which have been selected based on a review of the submitted abstracts.

The topic of the symposium was "100 Years ISPRS – Advancing Remote Sensing Science" to celebrate the foundation of the International Society for Photogrammetry (ISP) on July 4, 1910, on the initiative by Prof. Eduard Doležal. The Society changed its name to the International Society for Photogrammetry and Remote Sensing (ISPRS) in 1980, reflecting the increasing integration of the two disciplines. In our modern digital age, photogrammetry and remote sensing have virtually grown together, having as their common scope the extraction of reliable information from non-contact imaging and other sensor systems about the Earth and its environment through recording, measuring, analysing and representation.

Given the particular occasion the themes of the symposium extended beyond the traditional realms of Commission VII ("Thematic Processing, Modelling and Analysis of Remotely Sensed Data") by inviting contributions from the other ISPRS Commissions as well. Contributions that provided a comprehensive overview of the major research areas in remote sensing, highlighting past achievements and identifying challenges for the future, were particularly welcome. The conference topics were

- Multi-spectral and hyperspectral remote sensing
- Microwave remote sensing
- Lidar and laser scanning
- Geometric modelling
- Physical modelling and signatures
- Change detection and process modelling
- Land cover classification
- Image processing and pattern recognition
- Data fusion and data assimilation
- Earth observation programmes
- Remote sensing applications
- Operational remote sensing applications

We would like to take this opportunity here to sincerely thank the reviewers of both the full-papers and the abstracts for their valuable time and expertise! Each paper within Part A was reviewed by at least two reviewers and revised according to their comments. Only 60% of the submitted full papers finally made it through this review process. Also each abstract was reviewed by two peers, most of them by even three of more peers. Without question, the reviewer's effort was not in vain as it has helped to raise the quality of the papers and has allowed us to put together a high-quality technical programme.

Nevertheless, we also would like to add some self-criticism here. Within ISPRS the importance of a proper review process is broadly recognised. Also, more and more ISPRS colleagues rightfully request that all papers published in the *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (such as those contained in the current volume) shall be covered by prominent indexing and abstracting databases. Yet, there is no standard ISPRS reviewing system, nor is there a professional editorial support and indexing service. Consequently, every organiser of an ISPRS workshop, symposium or congress needs to reinvent the wheel by building up his/her own system for reviewing and handling the scientific contributions and for producing the proceedings. Having never organised an event with more

than hundred people, the learning curve has been particularly steep for us. As a result, mistakes were unavoidable and we would like to apologise with the authors and reviewers for any inconvenience that they may have experienced. We are confident that together we will be able to solve this challenge for future scientific meetings to the benefit of our discipline.

Finally, we want to highlight the very positive fact that we have received many excellent papers from scientists from developing regions. This suggests to us that ISPRS has been quite successful in stimulating the exchange of scientists from around the world and diverse cultural background. This will become the ever more important as climate change, continued population growth and shrinking natural resources have all become truly global problems that require, as one small part of the solution, global observation capabilities to better understand of how we have to act locally.

Vienna, June 2010

Wolfgang Wagner

Balázs Székely

Symposium Committee

Organising Committee

Wolfgang Wagner, Vienna University of Technology, Austria (Chairman)
Balázs Székely, Vienna University of Technology, Austria (Scientific Coordinator)

Wouter Dorigo, Vienna University of Technology, Austria
Michael Franzen, Bundesamt für Eich- und Vermessungswesen, Austria
Juha Hyyppä, Finnish Geodetic Institute, Finland
Josef Jansa, Vienna University of Technology, Austria
Norbert Pfeifer, Vienna University of Technology, Austria
Werner Schneider, University of Natural Resources and Applied Life Sciences, Austria
Klaus Steinnocher, Austrian Institute of Technology, Austria

Scientific Committee

Richard Bamler, DLR, Germany
Arnold Dekker, CSIRO, Australia
Diego Fernandez, ESA
Wolfgang Förstner, Bonn University, Germany
Garik Gutman, NASA, USA
Michael Schaepman, Universität Zürich, Switzerland
Johannes Schmetz, EUMETSAT
Uwe Sörgel, Leibniz Universität Hannover, Germany
Niko Verhoest, Ghent University, Belgium
Jixian Zhang, Chinese Academy of Surveying and Mapping, China
Yun Zhang, University of New Brunswick, Canada
Qiming Zhou, Hong Kong Baptist University, China

Reviewers

We gratefully acknowledge the support of the following colleagues who generously helped us in the review of the full-papers and abstracts:

Luis Carlos Alatorre
Paul Aplin
Roman Arbiol
Costas Armenakis
Clement Atzberger
Christoph Aubrecht
Manos Baltsavias
Richard Bamler
Georg Bareth
Zoltan Bartalis
Annett Bartsch
Enton Bedini
Eyal Ben-Dor
Santosh P. Bhandari
Horst Bischof
Christian Briese
Manfred Buchroithner
Ana Cabral
Liang-Chien Chen
Jan Clevers
Rene Colditz
Jackson Cothren
Jean-Francois Crétaux
Michele Crosetto
Wouter Dorigo
Peter Dorninger
Ian Dowman
Diego Fernandez
Michael Franzen
Steffen Fritz
Garik Gutman
Stefan Hasenauer
Christian Heipke
Manuela Hirschmugl
Thorsten Hoberg
Bernhard Höfle
Markus Hollaus
Pai-Hui Hsu
Tangao Hu
Juha Hyyppä
Arnold Irschara
Josef Jansa
Sanna Kaasalainen
Helmut Kager
Wilfried Karel
Akira Kato
Viktor Kaufmann
Akila Kemmouche
Richard Kidd
Géza Király
Mathias Kneubühler
Ilkka Korpela
Tatjana Koukal
Nelley Kovalevskaya
Olaf Kranz
Claudia Künzer
Peijun Li
Bo-Cheng Lin
Roderik Lindenbergh
Zhengjun Liu
Zhong Lu
Aliaksei Makarau
Clément Mallet
Gottfried Mandlburger
Lauri Markelin
Helmut Mayer
Ian McCallum
Thomas Melzer
Gábor Molnár
Werner Mücke
Sonoyo Mukai
Vahid Naeimi
Alexey Nekrasov
Wenige Ni-Meister
Johannes Otepka
Gintautas Palubinskas
Nicolas Paparoditis
Charles Paradzayi
Sang-Eun Park
Carsten Pathe
Norbert Pfeifer
Camillo Ressl
Victor F. Rodríguez-Galiano
Petri N. Rönnholm
Dominik Rosenbaum
Franz Rottensteiner
Nilda Sánchez
Michael Schaepman
Johannes Schmetz
Thomas Schneider
Alexander Schunert
Gabriela Seiz
Jie Shan
George Sithole
Julian Smit
Ben Somers
Uwe Sörgel
Klaus Steinnocher
Uwe Stilla
Haigang Sui
Anna Nora Tassetti
Tee-Ann Teo
Thi Nhat Thanh Nguyen
Gábor Timár
John Trinder
Jaan-Rong Tsay
Niko Verhoest
Loureal Camille Verzosa
Karsten Vogt
Francesco Vuolo
Wolfgang Wagner
Peter Waldhäusl
Miao Wang
Björn Waske
Aloysius Wehr
Uwe Weidner
Gui-Song Xia
Mitsunori Yoshimura
Yu Zeng
Jixian Zhang
Li Zhang
Yun Zhang
Qiming Zhou

THE EFFECT OF BIOMASS AND SCANNING ANGLE ON LASER BEAM TRANSMITTANCE

E. Ahokas [a, *], J. Hyyppä [a], H. Kaartinen [a], A. Kukko [a], S. Kaasalainen [a], A. Krooks [a]

[a] Finnish Geodetic Institute, Geodeetinrinne 2, 02430 Masala, Finland
eero.ahokas@fgi.fi

KEY WORDS: Laser scanning, point cloud, DEM/DTM, accuracy, experiment

ABSTRACT:

During the last decade, there have been numerous scientific studies verifying the accuracy of digital elevation models (DEM) derived from airborne laser scanning (ALS). Since ALS has increasingly been used for nationwide digital elevation model data acquisition, optimizing ALS acquisition parameters is a topic of interest to national land surveys. In particular, the effect of the scanning angle and biomass on elevation-model accuracy needs further study in heavily-forested areas. The elevation-model accuracy is affected by, for example, the number of pulses hitting the ground, footprint size, terrain slope and, especially, vegetation. In order to better understand the effect of the biomass and scanning angle on the penetration rate of ALS signal through canopy and give further support to ALS studies, especially for scanning angles beyond 15 degrees of the nadir point, we conducted an indoor experiment using small spruce trees to represent forest canopy. The indoor experiment allowed us to measure the biomass reference accurately. We used manual thinning to produce various levels of biomass and scissor lift as the carrying platform. We measured the weight of every tree and the total biomass of trees after each thinning phase. We removed the material homogeneously from the trees, starting from the latest shoots. We used a FARO laser scanner in the experiment and attached it to the scissor lift. We scanned the experimental plot from four altitudes (about 3, 5, 7 and 9 m) and at six biomass levels (about 0, 6, 9, 14, 20 and 25 kg). The results show that signal transmittance through spruce trees is a function of biomass and scanning angle, but that the scanning angle only has a minor effect on the results. Biomass is the major parameter in determining the quality of the elevation model. While the results require further airborne experiments to be fully confirmed, they do imply that a scanning angle greater than 15 degrees can be applied in regions having low and moderate biomass, and due to the significant effect of the biomass on the transmittance, the airborne scanning missions must be carefully specified in heavily-forested terrain. We also found that terrestrial laser scanning experiments performed in an indoor laboratory-type setting yielded a relatively good understanding of the basic behaviour of and interaction between the target and laser scanning rather easily, but that it will be considerably more difficult to obtain similar results in a real-life experiment due to limited accuracy when collecting the reference data.

1. INTRODUCTION

Airborne laser scanning has been used for topographic mapping and forestry applications for many years. The accuracy of Digital Elevation Models obtained over forested areas has been described by, for example, Kraus and Pfeifer (1998), Hyyppä et al. 2001), Ahokas et al. (2002), Reutebuch et al. (2003), Takeda (2004), Sithole and Vosselman (2004), Hyyppä et al. (2005), Su et al. (2006), Chasmer et al. (2006), and Morsdorf et al. (2008). A detailed comparison of the filtering techniques used for DEM extraction was made within an ISPRS comparison of filters (Sithole and Vosselman 2004). Reutebuch et al. (2003) reported random errors of 14 cm for clear-cuts, 14 cm for heavily-thinned forest, 18 cm for lightly-thinned forest and 29 cm for uncut forest, using TopEye data with 4 pulses per m^2. The variation in ALS-derived DEM quality with respect to date, flight altitude, pulse mode, terrain slope, forest cover and within-plot variation was reported by Hyyppä et al. (2005).

Ahokas et al. (2005) proposed that the optimization of the scanning angle (i.e. field of view) is an important part of nationwide laser scanning. Significant savings can be realized by increasing the scanning angle and flight altitude. The initial results obtained using scanning angle analysis showed that the scanning angle had an effect on the accuracy of Digital Elevation Models, but that other factors, such as forest density, dominate the process. Scanning angles up to 15 degrees seem to be usable for high-altitude laser scanning in the boreal forest zone. High-altitude laser scanning yielded a precision measurement of about ±20 cm (std), which is good enough for most terrain models required in forested areas. Ahokas et al. (2005) stressed that the effects of the scanning angle should be studied further, since the maximum field of view for commercial laser scanners can be up to 75 degrees.

Su et al. (2006) analyzed the influence of vegetation, slope and the LiDAR sampling angle (the laser beam angle from nadir) on DEM accuracy. Vegetation caused the greatest source of error in the LiDAR-derived elevation model. It was also reported that DEM accuracy decreased when the slope gradient increased. Off-nadir scanning angles should be less than 15 degrees to minimize the errors coming from high slope gradients. The LiDAR sampling angle had little impact on the measured error. Chasmer et al. (2006) investigated laser pulse penetration through a conifer canopy by integrating airborne and terrestrial LiDAR. They found that pulses with higher energy penetrate further into the canopy. The authors suggest that future research should concentrate on improving the understanding of how laser-pulse returns are triggered within vegetated environments and how canopy properties influence the location of the trigger event. Morsdorf et al. (2008) assessed the influence of flying altitude and scanning angle on biophysical vegetation products (tree height, crown width, fractional cover and leaf area index) derived from airborne laser scanning. Due to the small scanning angle of the TopoSys Falcon II (±7.15 degrees), the dependence of airborne laser scanning on the incidence angle is not so evident. The incidence angle (angle to surface normal of the horizontal plane) seems to be of greater importance for vegetation density parameters than the local incidence angle (the angle to surface normal in the elevation model). The local

* Corresponding author

topography is thus less important than the scanning angle. ALS data from larger scanning angles should be used to study further the effect of scanning angle on vegetation density products.
In order to separate the effects of vegetation and the scanning angle on the accuracy of the elevation model, especially beyond 15 degrees of the nadir point (which has not yet been studied), we carried out an indoor experiment. The controlled indoor conditions eliminated the effect of slope from the possible sources of error present in terrain measurements. In this test, we took the number of laser beams that were emitted versus those that were recorded hitting the ground and compared them using various canopy densities and scanning angles. The basic assumption of the study was that the number of laser hits reaching the ground serves as the main source affecting elevation model quality. Thus, we studied the transmittance through the canopy as the function of biomass and scanning angle. The laser beam size and triggering algorithm that we applied differs from system to system and from terrestrial to airborne systems, and, thus, the final results need to be verified in future airborne tests.

2. MEASUREMENTS

2.1 Terrestrial laser scanning

The applied terrestrial laser scanner (TLS) was a FARO Photon 80, which is based on phase measurements that provide high-speed data acquisition. The technical parameters of the scanner include a maximum measurement rate of 120,000 points/s, a wavelength of 785 nm, a vertical field of view of 320 degrees and a horizontal field of view of 360 degrees, and a ±2 mm systematic distance error at 25 m. A beam divergence is 0.16 mrad and a beam diameter at exit is 3.3 mm (circular) (www.faro.com). The resolution setting that we used was 1/8 of the full scanning resolution, which is to say 13.9 points/degree or 0.072 degrees/point. The resolution is the same for the vertical and horizontal directions. The phase-shift-based system uses an amplitude-modulated laser beam. The phase shift between the transmitted and the received signal is measured. Ambiguity can be resolved by using the multi-wavelength phase shifts (RP Photonics 2010, Kikuta et al. 1986). If the continuous, multi-wavelength amplitude-modulated beam hits multiple targets, the measurement range is not well-defined. The ALS is typically a pulse-based system which gives multiple returns. The phase-shift-based system can be used to approximate the penetration rate of the ALS pulse-based system, since ground return does not exist when several hits are encountered. Thus, the phase-shift-based system gives a lower bound (worst case scenario) for the penetration rate than the pulse-based system does. Since the triggering algorithm of each laser scanner is different, and is also affected by the laser beam size and the sensitivity of the receiver, the results need to be verified separately for airborne systems. The results can be used to better understand the effect of the scanning angle in relation to the biomass on elevation models.

For this study, we used a Genie GS3232 scissor lift, which has a maximum platform level of 9.8 m. We attached the FARO scanner to the front safety fence of the platform. The level and the function of the scanner were controlled from the lobby floor. Cf. Figures 1 and 2.

We scanned the experimental plot from four altitudes (about 3, 5, 7 and 9 m) and at six biomass phases (about 0, 6, 9, 14, 20 and 25 kg (Fig. 3)). The scanning angle is the angle between the nadir point under the laser scanner, the laser scanner and the point at which the laser beam hits the target. By changing the scanning height, it was possible to record a larger scale of scanning angles (from 6 to 38 degrees) over the test plot than by using only one scanning height. The scanning angle categories for 9, 7, 5 and 3 m heights were 6-15, 8-19, 11-26 and 17-38 degrees (Table 6).

Figure 1. Spruce trees and the FARO scanner in the lobby. Thinning phase th0. Photo M. Kurkela.

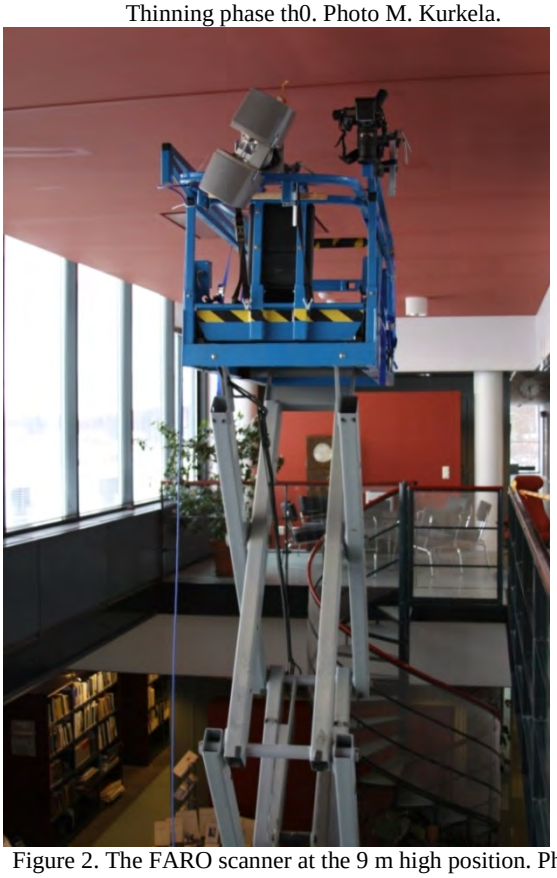


Figure 2. The FARO scanner at the 9 m high position. Photo J. Hyyppä.

2.2 Reference data

We placed the spruce trees (*Picea abies*) on a base (base height about 20 cm) on the floor of the Institute lobby at a point that was high enough for us to take measurements from different altitudes using a lift. We weighed every tree before commencing with the experiment. We extracted the biomass from the trees gradually by hand. We also weighed the thinned parts of the trees (mainly branches) after every thinning phase. We removed the latest shoots first, after which we removed the rest of the parts of the trees at an even rate. Figure 3 shows the decreasing amount of the total biomass. Our plan for cutting an equal amount of mass at every phase was quite successful, which is indicated by the almost linear curve for the diminishing biomass.

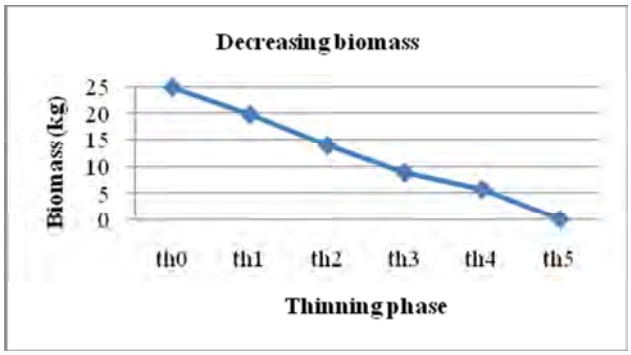


Figure 3. Spruce biomass decreasing during the experiment.

The thinning phases can be described as follows: Th0 indicates the untouched original biomass, when all the needles and branches were still on the trees (Fig. 1); during phases th1 and th2 some of the branches were removed (Fig. 4); during phase th3 only minor branches and the tree trunks remained; th4 indicates a clear-cut situation and the final phase, th5, indicates the cleaned plot area, when all the laser beams hit the ground without encountering any obstacles.

Figure 4. Thinning phase th2.

2.3 Analysis

For the statistical analysis, we used the two-factor analysis of variance. In the relevant literature, it is also called the complete block design (Montgomery 1984). With this analysis it is possible to find out whether the two factors differ from each other. Test statistic F is defined as

$$F = \frac{MS_{factor}}{MS_{err}}, \quad (1)$$

where MS_{Factor}=mean square of the factor
MS_{err}= mean square of the error

We compared test statistic F with the F distribution critical value $Fcrit_{\alpha,\upsilon1,\upsilon2}$ to determine its significance at the significance level of α=0.05. $\upsilon1$ represents the degree of freedom of the factor and $\upsilon2$ the degree of freedom of the error. If $F>Fcrit_{\alpha,\upsilon1,\upsilon2}$, then there are statistically significant differences within the factor. In this study, the biomass of the trees was the first factor and the scanning angle was the second factor. In Tables 7, 8, 10 and 12, P stands for the probability that the result for the significance was purely a coincidence.

We determined the laser points transmittance as a ratio of the ground points (H<20 cm above the floor) and the total number of points in the plot. Using this height limit, we considered the tree base points as ground points.
To study the effect of the scanning angle, we divided the entire view of interest at each scanning height into three parts. For example, we stratified scanning angles between 6 and 15

degrees at the 9 m scanning height into 6-8, 9-11 and 12-15 degree categories. In a similar manner, we stratified the observations from other scanning heights into three categories. The mean values of each scanning angle category are presented in Fig. 6 and 7. We examined the F statistics using all the scanning angle categories (12=4 heights x 3 categories) within the same scanning angle factor.

We analyzed the front edge of the test plot facing the scanner side separately. In this region, the spruce trees have fewer branches that prevent the laser beam from penetrating the canopy. We have studied the effects from this phase of the research by analysing the so-called front part angles separately.

3. RESULTS

Laser beams penetrate the original dense spruce stand poorly; (Figure 5 and Table 6) their rate of transmittance to the ground level is about 1-2%. After the thinning at phase th1, the transmittance rate was 5-6%. The next thinning phase increased the transmittance rate to 20-31%. The giant leap in the amount of transmittance after that is the result of leaf area loss when the branches are cut off.

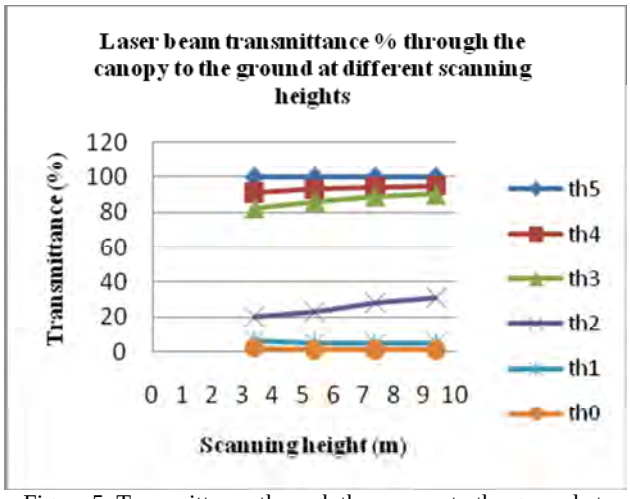


Figure 5. Transmittance through the canopy to the ground at different scanning heights and thinning phases.

Biomass (kg), thinning phase	T(%), 9 m, 6˚-15˚	T(%), 7 m, 8˚-19˚	T(%), 5 m, 11˚-26˚	T(%), 3 m, 17˚-38˚
24.84 (th0)	1	1	1	2
19.76 (th1)	5	5	5	6
13.88 (th2)	31	28	23	20
8.76 (th3)	90	89	86	82
5.68 (th4)	95	94	93	91
0 (th5)	100	100	100	100

Table 6. Transmittance T as a percentage of the laser beams reaching the ground from four altitudes. Minimum and maximum scanning angles in degrees.

When we treated each of the scanning heights as a single-angle category (data from Table 6), the biomass served as a significant factor in transmittance change at the α=0.05 significance level whereas the scanning angle did not (Table 7).

Factor	F	P	Fcrit	
Biomass	1635.55	5.73E-20	2.90	*
Scanning angle	2.79	0.08	3.29	

Table 7. F statistic for one scanning angle category per scanning height. Significance level α=0.05. *=factor statistically significant.

When we divided the scanning angles into three equal parts at each scanning height, the scanning angle factor contains a total of 12 angle categories to study instead of only 4, as in Table 7. Table 8 also shows that the scanning angle now becomes a significant factor affecting the rate of transmittance through the spruce canopy.
We also examined the three angle categories per each height separately. The F statistic is presented in Table 10.

Factor	F	P	Fcrit	
Biomass	668.43	6.43E-48	2.38	*
Scanning angle	4.76	4.2E-05	1.97	*

Table 8. F statistic for all scanning angles (12 categories). Significance level α=0.05. *=factor statistically significant.

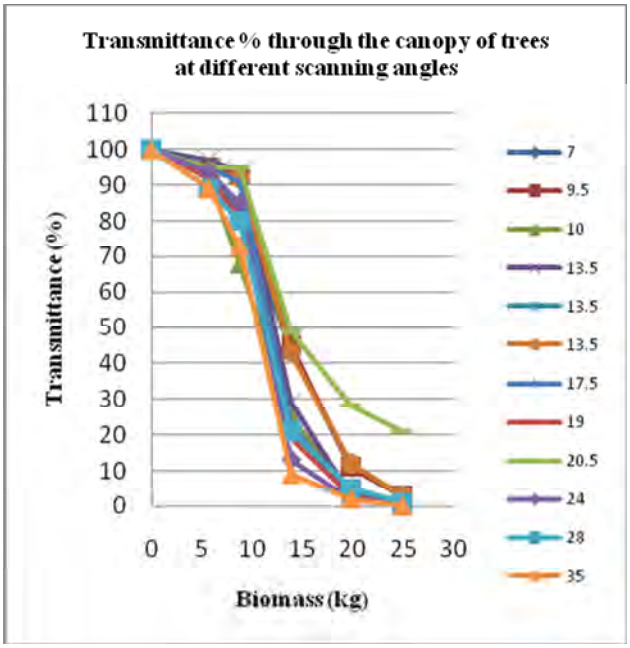


Figure 9. Transmittance through the canopy of trees at different scanning angles (from 7 to 35 degrees). We stratified the scanning angles into three parts at each height. The angle value presented is the mean value of each part.

Height (m)	Factor	F	P	Fcrit	
3	Biomass	95.08	4.24E-08	3.33	*
	Scanning angle	10.82	3.15E-03	4.10	*
5	Biomass	173.38	2.23E-09	3.33	*
	Scanning angle	3.63	0.07	4.10	
7	Biomass	249.01	3.73E-10	3.33	*
	Scanning angle	3.45	0.07	4.10	
9	Biomass	148.79	4.74E-09	3.33	*
	Scanning angle	2.34	0.15	4.10	

Table 10. F statistic for every scanning height when scanning angles were divided into three parts. Significance level α=0.05. *=factor statistically significant.

The scanning angle was only significant during observations at the scanning height of 3 m at the α=0.05 significance level. At that height, the variability in scanning angles was 17-38 degrees, making it significantly larger than at other heights. Furthermore, the biomass was a significant factor at all scanning heights.

The laser beam transmittances through the canopy and the scanning angles in front of the test plot facing the scanner are depicted in Figure 11. The scan (mean angle=20.5 degrees) taken at the height of 3 m differs from the other scans.

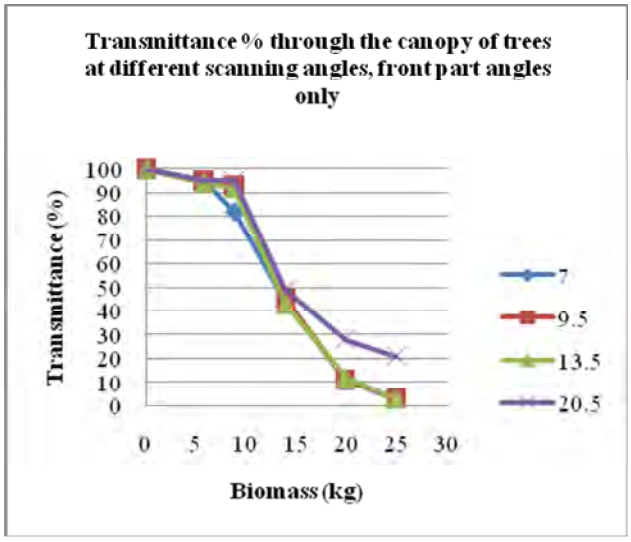


Figure 11. Transmittance through the canopy of trees at different scanning angles. Only the front part angles are presented for each scanning height (9, 7, 5 and 3 m).

If we omit the front part angles (with mean values of 7, 9.5, 13.5, and 20.5 degrees) from the scanning angle factor and examine the F statistics using all the remaining scanning angles in the angle factor, then the scanning angles are no longer statistically significant at the α=0.05 significance level. A significantly larger test plot would have helped to eliminate this effect caused by the front part angles.

Factor	F	P	Fcrit	
Biomass	990.39	1.23E-36	2.49	*
Scanning angle	2.11	0.07	2.29	

Table 12. F statistic when front part scanning angles were extracted. Significance level α=0.05. *=factor statistically significant.

4. DISCUSSION

The scanning angle had only a minor effect on the results compared to changes in the biomass. Dense canopy was the main source of transmittance deterioration as well as elevation-model accuracy deterioration. From the point of view of specifying ALS DEM quality, the results confirm the initial assumptions by Hyyppä et al. (2005) and Su et al. (2006) that the canopy is the main source for errors in the DEM accuracy. In order to be able to specify DEM quality, one has to know the properties of the forest canopy or be able to accurately specify the necessary accuracy for certain biomass or forest-condition levels. Thus, it is not feasible to specify acceptable rates for DEM accuracy without knowing the forest biomass.

In this study, we used a terrestrial laser scanner instead of an airborne scanner, since we needed to achieve an accurate reference. If there are multiple reflections from the target area, the distance given by the phase measurement method is not well-defined. Thus, this system exaggerates the effect of the scanning angle and diminishes the amount of ground-reflected points. Since the transmittance rate through the forest can be only a matter of a few percentage points, even a reasonable increase in the number of transmitted laser beams does not necessarily help in reaching acceptable DEM accuracy rates. We, therefore, believe that the results give valuable guidelines for pulse-based ALS, even though the beam size and different triggering algorithm varies from system to system, meaning that the results need to be verified through airborne experiments. Based on this experiment as well as prior research on forested conditions, we recommend using a number of pulses per square metre as a feasible criterion in specifying the laser scanning missions and making offers comparable rather than trying to obtain sufficient elevation model accuracy by specification, which requires a priori knowledge of forest conditions.

The authors believe that the most valuable contribution of this study is that it demonstrates how laboratory-type indoor terrestrial laser scanning experiments can be used to study the basic behaviour of the target and laser scanning interaction, which would be far too complicated to carry out via commercial ALS experiments. In this case, the number of sensors providing large scanning angles was limited (consisting only of Leica sensors), and it was very difficult to carry out enough accurate biomass measurements for all scan angles and biomass classes. Therefore, we urge researchers to conduct more laboratory-type studies in the future, which will provide better understanding of the basic interaction of laser beams with the target. Within the scientific community, acceptance of ALS has proceeded at a much faster rate than supporting research. Therefore, small laboratory-type experiments, such as the one depicted in this paper, can provide a quick and basic understanding of the phenomena with lower costs. The final conclusion can then be more easily confirmed using airborne experiments.

5. CONCLUSION

Results show that laser-beam transmittance through a small canopy of spruce trees is a non-linear function of biomass. The scanning angle has only a minor effect on the results compared to changes in the biomass. Scanning angles up to 38 degrees proved feasible for elevation mapping through this indoor experiment. Thus, airborne experiments which have a scanning angle greater than 15 degrees still need to be performed, especially in areas with low and moderate levels of biomass. Dense canopy was the main source of transmittance deterioration and, thus, of elevation model accuracy deterioration.

We showed by way of a light experiment that laboratory-type indoor terrestrial laser scanning experiments can be used to study the basic behaviour of and interaction between the target and laser scanning, which confirms previous results from airborne experiments and suggests new possibilities for extending the scanning angle for ALS surveys of areas with low and moderate levels of vegetation.

REFERENCES

Ahokas, E., Kaartinen, H., Matikainen, L., Hyyppä, J. and Hyyppä, H., 2002, Accuracy of high-pulse-rate laser scanners for digital target models. In: Begni, G. (ed) *Observing our environment from space. New solutions for a new millennium, Proceedings of the 21st EARSeL Symposium, 14-16 May 2001*, (Paris: Balkema Publishers), pp. 175-178.

Ahokas, E., Yu, X., Oksanen, J., Hyyppä, J., Kaartinen, H.,Hyyppä, H. 2005. Optimization of the scanning angle for countrywide laser scanning. In: Vosselman, G., Brenner, C. (eds.) ISPRS Workshop Laser scanning 2005. Enschede, the Netherlands 12-14 September 2005. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVI, part 3/W19, pp. 115-119, ISSN 1682-1750.

Chasmer, L,. Hopkinson, C., Treitz, P., 2006. Investigating laser pulse penetration through a conifer canopy by integrating airborne and terrestrial lidar. *Canadian Journal of Remote Sensing*. Vol. 32, No 2, pp116-125.

Hyyppä, J., Schardt, M., Haggrén, H., Koch, B., Lohr, U., Scherrer, H.U., Paananen, R., Luukkonen, H., Ziegler, M., Hyyppä, H., Pyysalo, U., Friedländer, H., Uuttera, J., Wagner, S., Inkinen, M., Wimmer, A., Kukko, A., Ahokas, E. and Karjalainen, M., 2001, HIGH-SCAN: The first European-wide attempt to derive single-tree information from laserscanner data. *The Photogrammetric Journal of Finland*, Vol.17, pp. 58-68.

Hyyppä, H., Yu, X., Hyyppä, J., Kaartinen, H., Kaasalainen, S., Honkavaara, E., Rönnholm, P., 2005. Factors affecting the quality of DTM generation in forested areas. In: Vosselman, G., Brenner, C. (eds.) ISPRS Workshop Laser scanning 2005. Enschede, the Netherlands 12-14 September 2005. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVI, part 3/W19. pp. 85-90. ISSN 1682-1750.

Kikuta, H., Iwata, K. and R. Nagata, 1986. Distance measurement by the wavelength shift of laser diode light. *Applied Optics*, 25 (17), 2976- .

Kraus, K., Pfeifer, N., 1998. Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 53 No 4, pp. 193-203.

Montgomery, D., 1984. *Design and analysis of experiments.* John Wiley and Sons, ISBN 0-471-88908-3, pp. 123-130.

Morsdorf, F., Frey, O., Meier, E., Itten, K., Allgöwer B., 2008. Assessment of the influence of flying altitude and scan angle on biophysical vegetation products derived from airborne laser scanning. *International Journal of Remote Sensing*. Vol. 29, No 5, pp. 1387-1406.

Reutebuch, S., McGaughey, R., Andersen, H-E., Carson, W., 2003. Accuracy of a high-resolution lidar terrain model under a conifer forest canopy. *Canadian Journal of Remote Sensing*. Vol. 29, No. 5, pp. 527-535.

RP Photonics, 2010. *Encyclopedia of laser physics and technology*.
URL:http://www.rpphotonics.com/phase_shift_method_for_distance_measurements.html. Accessed 12.2.2010.

Sithole, G,. and Vosselman, G., 2004. Experimental comparison of filter algorithms for bare Earth extraction from airborne laser scanning point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*. Vol. 59 (1-2), Pp. 85-101.

Su, J., Bork, E., 2006. Influence of vegetation, slope and Lidar sampling angle on DEM accuracy. *Photogrammetric Engineering and Remote Sensing*. Vol. 72, No. 11, pp. 1265-1274.

Takeda, H., 2004, Ground surface estimation in dense forest. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol.35, (part B3), pp. 1016-1023.

URL:http://www.faro.com

ANALYSIS OF SPATIAL AND TEMPORAL EVOLUTION OF THE NDVI ON VEGETATED AND DEGRADED AREAS IN THE CENTRAL SPANISH PYRENEES

L. C. Alatorre [a,] *, S. Beguería [b]

[a] Pyrenean Institute of Ecology, CSIC, Campus de Aula Dei, Apdo 202, 50080 Zaragoza (Spain) - lalatorre@ipe.csic.es
[b] Estación Experimental de Aula Dei, CSIC, Av. Montañana, 1005, 50059, Zaragoza (Spain) - sbegueria@eead.csic.es

KEY WORDS: Algorithms, Specification, Spatial, Vegetation, Land Cover

ABSTRACT:

The temporal evolution of vegetation activity on various land cover classes in the Spanish Pyrenees was analyzed. Two time series of the normalized difference vegetation index (NDVI) were used, corresponding to March (early spring) and August (the end of summer). The series were generated from Landsat TM and Landsat ETM+ images for the period 1984-2007. An increase in the NDVI in March was found for vegetated areas, and the opposite trend was found in both March and August for degraded areas (badlands and erosion risk areas). The rise in minimum temperature during the study period appears to be the most important factor explaining the increased NDVI in the vegetated areas. In degraded areas, no climatic or topographic variable was associated with the negative trend in the NDVI, which may be related to erosion processes taking place in these regions.

1. INTRODUCTION

Maps of active erosion areas and areas at risk of erosion are of great potential use to environmental agencies (governmental and private), as such maps allow erosion prevention efforts to be concentrated in places where the benefit will be greatest. There is no single straightforward method for assessing erosion, as erosion is highly dependent on the spatial scale and the purpose of the assessment (Warren, 2002). Methods for evaluating erosion risk at catchment and regional scales (10-10,000 km^2) include the application of erosion models, or qualitative approximations using remote sensing (RS) and geographic information system (GIS) technologies. RS and GIS techniques have been shown to be of potential use in erosion assessment at regional scales, including the identification of eroded surfaces, estimation of factors that control erosion, monitoring the advance of erosion over time, and investigating vegetation characteristics and dynamics (Lambin, 1996).

Various studies have identified changes in vegetation dynamics at continental, regional, and local scales in recent decades. Most changes have been caused by human activity, particularly deforestation and forest fires (Riaño et al., 2007), but land marginalization and rural abandonment have contributed to natural revegetation processes in some regions (Vicente-Serrano et al., 2004). However, numerous reports have found a general increase in vegetation activity in various ecosystems of the world, suggesting that the principal causes of changes in vegetation dynamics are variations in precipitation and/or temperature (Delbart et al., 2008).

Changes in vegetation in the Mediterranean region have followed very different patterns. In general, vegetation growth tends to be favored by increased temperature in areas where water is not a limiting factor (Martínez-Villalta et al., 2008). Studies in the Spanish Pyrenees (Lasanta and Vicente-Serrano, 2007) have investigated spatial and temporal variations in vegetation cover at regional and local sc ales to i) assess changes in the vegetal cover in the last 50 years; ii) detect trends in the global vegetation biomass; iii) explore changes in leaf activity in forest regions; iv) detect the climate drivers (temperature and precipitation) and spatial patterns of aridity; and, v) to determine the effects of human land uses. All previous studies have focused on well-vegetated areas, and very few reports have analyzed spatial and temporal variations in vegetation cover on active erosion areas and erosion risk areas, where vegetation is sparse. Badlands are usually defined as intensely dissected natural landscapes where vegetation is scanty or absent. Alatorre and Beguería (2009) identified active erosion and erosion risk areas in a badlands landscape of the Spanish Pyrenees using RS techniques. The presence of bare soil surfaces and the large size of badlands enabled good discrimination using RS data. However, the erosion risk areas surrounding badlands, coinciding with the transition zone from badlands to scrubland or forest, were characterized by poor vegetation cover (10-50%). For this reason, the analysis of vegetation dynamics on active erosion and erosion risk areas is very relevant to the design of measures for the mitigation and remediation of soil erosion and sediment transfer.

The objectives of this study were i) to obtain time series of vegetation activity during two contrasting periods of the growth cycle (early spring and the end of summer) for various land cover classes, including both well-vegetated and degraded areas; ii) to determine the extent by which climate controlled vegetation activity in the various land cover classes, and to define temporal trends; and, iii) to analyze the spatial distribution of trends in vegetation activity on erosion risk areas, as indicators of recovery and degradation, and to quantify the effects of various topographical factors on such trends.

2. STUDY AREA

The study area, located at 620-2,149 m altitude approximately 23 km north of the Barasona Reservoir (Spanish Pyrenees), is an integrated badlands landscape orientated northwest−southeast and developed on Eocene marls (Fig. 1A and B). A land cover map based on the supervised maximum likelihood method (Alatorre and Beguería, 2009) showed that the study area is occupied by five principal land cover categories: badlands, 19 km^2 (8.0%); coniferous forest, 65 km^2 (28.0%); deciduous forest, 21 km^2 (9.0%); grassland, 32 km^2

* Corresponding author

(13.0%); and scrubland, 99 km^2 (42.0%). The spatial distribution of land cover showed that the areas occupied by scrub and the grass border areas could be classified as badlands (Fig. 1C). This spatial distribution suggested that a progressive transition between eroded areas and forest (Fig. 1C). In the same study, a map of the active erosion (badlands) and erosion risk areas was obtained, with the surface areas of these classes comprising 17 km^2 and 49 km^2, respectively (Fig. 1D). The surface area of the active erosion region was the same as that obtained from a land cover map generated using the supervised maximum likelihood method. The badlands system comprises a group of typical hillside badlands developed on sandy marls with clay soil, and is strongly eroded over convex hillsides with moderately inclined slopes. Visual comparison of maps showed that the erosion risk areas corresponded principally to the scrubland class (and in some cases the grassland and conifer classes) bordering the badland areas. These areas had spectral characteristics intermediate between badlands and scrubland, indicating either a mixture of classes within a pixel or an intermediate level of degradation (for more details of the study area please see, Alatorre and Beguería, 2009).

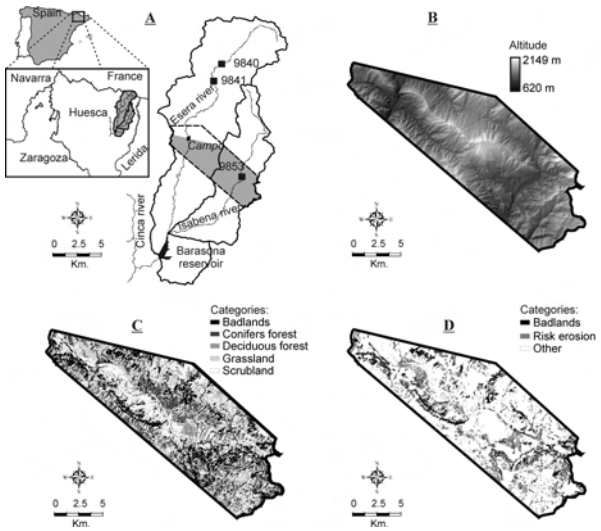


Figure 1. A) Location of the study area: i) subset area indicates the location of badland areas on marls (236 km^2); ii) the gray zone indicates the area of the Landsat scene; iii) the black squares indicate the location of meteorological observatories of the National Agency of Meteorology. B) Digital terrain model (DTM). C) Land cover map based on supervised classification using the maximum likelihood method and the maximum probability classification rule (Alatorre and Beguería, 2009). D) Erosion risk maps (Alatorre and Beguería, 2009).

3. DATA AND METHODS

3.1 Data selection and preparation

A database of Landsat TM and Landsat ETM+ images for the period 1984-2007 was used. The database comprised 28 images, 16 of which were from a summer time series and 12 from a spring time series. The two time series were used to identify possible differences in vegetation dynamics as a function of seasonal differences in vegetation activity, and to assess with more robustness any spatial and temporal patterns in vegetation activity. Table 2 shows the dates of the images used in each time series. The database was processed using a procedure that included calibration and cross calibration of the images (for more details please see, Alatorre and Beguería, 2009). The procedure allowed accurate measurements of physical surface reflectance units to be obtained. The correction applied to the images guaranteed the temporal homogeneity of the dataset, the absence of artificial noise caused by sensor degradation and atmospheric conditions, and spatial comparability among different areas, given the accurate topographic normalization applied. Details of the correction procedure applied to the images, and a complete description of the dataset and its validation have been described by Vicente-Serrano et al. (2008).

Time series of the normalized difference vegetation index (NDVI) were obtained from the original Landsat TM and Landsat ETM+ images, for the purpose of monitoring vegetation activity. The NDVI was computed as (Rouse et al., 1974):

$$NDVI = \frac{\rho_{IR} - \rho_R}{\rho_{IR} + \rho_R} \qquad (1)$$

where ρIR is the reflectivity in the near-infrared region of the electromagnetic spectrum and ρR is the reflectivity in the red region. Several studies have demonstrated a strong relationship of the NDVI to the fraction of photosynthetically active radiation, the vegetation biomass, the green cover, and the leaf area index. Hence, high NDVI values are indicative of high vegetation activity. A land cover map comprising the major vegetation types in the study area was also used, as well as a map of active erosion areas (badlands) and areas at erosion risk (Alatorre and Beguería, 2009).

March		August	
Acquisition date	*Sensor*	*Acquisition date*	*Sensor*
03/11/1989	TM	08/20/1984	TM
03/30/1990	TM	08/07/1985	TM
03/06/1993	TM	08/13/1987	TM
03/09/1994	TM	08/02/1989	TM
03/28/1995	TM	08/24/1991	TM
03/17/1997	TM	08/10/1992	TM
03/20/1998	TM	08/29/1993	TM
03/23/1999	TM	08/03/1995	TM
03/17/2000	ETM+	08/24/1997	TM
03/10/2003	ETM+	08/14/1999	TM
03/07/2005	TM	08/08/2000	ETM+
03/13/2007	TM	08/26/2001	ETM+
		08/30/2002	ETM+
		08/27/2004	TM
		08/18/2005	TM
		08/01/2006	TM

Table 2. Dates for the Landsat 5 TM and 7 ETM+ images used in the study.

To analyze climate effects on the vegetation activity we used a database consisting of three daily rainfall series from the National Agency of Meteorology, comprising data since January 1984 (Fig. 1A). To guarantee the quality of the dataset the series were checked using a quality control process that identified anomalous records and analyzed the homogeneity of each series (for more details see Vicente-Serrano et al., 2009). Daily temperature data were obtained for the same period from the Serraduy station (Fig. 1A), and these were also checked for possible temporal inhomogeneities. The time series of precipitation totals and maximum/minimum temperature averages were computed from the original daily series by aggregating the original daily values over the period immediately before the images were taken. Thus, climatological series were computed for the following time periods prior to the date of the image: 15 days, 30 days, 3 months (January, February and March for the March images; June, July and August for the August images) and 6 months (October to March, and March to August, respectively). A series of

topographical variables was also analyzed to assess their effects on vegetation activity. This involved use of a digital terrain model (DTM) with a spatial resolution of 20 m to derive the slope gradient (m m^{-1}), as some studies have shown that this can be a major factor explaining rates of vegetation recovery (Pueyo and Beguería 2007). We also derived a model of the incoming solar radiation (MJ m^{-2} day^{-1}) to assess topographic control of the energy balance, using an algorithm that includes the effect of terrain complexity (shadowing and reflection) and the daily solar position (Pons and Ninyerola, 2008).

3.2 Statistical analysis

The temporal series of NDVI for each land cover class was checked for temporal trends using the Spearman's correlation test against time. This enabled analysis of the vegetation dynamics in terms of increased (positive correlation) and decreased activity (negative correlation). The significance of the trends was checked using the p value associated with the Spearman's rho statistic.

The Spearman's test enables detection of temporal trends in the NDVI series, but does not identify the driving factors involved. To determine the control exerted on vegetation activity by climate, and to isolate climate from other factors, we performed a multivariate regression analysis of the average NDVI values in March and August for the various land cover classes against the climatic variables. As a preliminary step we undertook a correlation analysis to determine the most appropriate time span for the climatologic time series. For both the March and August images we found that the climatological series computed for the 3 months prior to the images had the greatest correlation with the NDVI. Therefore, we used the time series of cumulative precipitation and average maximum/minimum temperature for the 3 months before the acquisition date as covariates in the regression analysis.
As the acquisition date of the images did not coincide among years, which could have affected the NDVI (especially in March, which is very close to the start of the growing period), we also introduced the Julian day of the image as a covariate. To check for temporal trends in the NDVI values that were not explained by variability of the climatic factors and the acquisition date of the images, we also incorporated the year of acquisition of the image as a covariate.
We used a backward stepwise procedure based on the Akaike's information criterion statistic (AIC), as implemented in the function stepAIC in the R package for statistical analysis (R Development Core Team, 2008). This function aided identification of the significant explanatory variables for the time evolution of the NDVI for the various land cover classes. The data analysis was based on the goodness of fit and statistical significance of the regressions, the explanatory variables selected, and the beta (standardized) regression coefficients.
To provide a spatially distributed analysis, the multivariate regression analysis was repeated on a pixel-by-pixel basis for the erosion risk areas alone. This enabled mapping of the spatial distribution of NDVI trends not explained by climatologic factors, and thus identification of areas undergoing processes of degradation or recovery. Finally, a correlation analysis was performed on the NDVI trends against various topographical factors (elevation, slope gradient and potential incoming solar radiation), and a bootstrap procedure was used to determine the statistical significance of the correlations. Thus, 1,000 repetitions of the correlation analysis were performed on random samples containing approximately 1% of the pixels belonging to the erosion risk class, and the resulting significance statistics (p values) were averaged. This enabled avoidance of a sample size effect that would arise if all the pixels of the erosion risk class (approximately 45,000) were introduced together in the analysis, causing the significance test to become over sensitive and thus unreliable.

4. RESULTS AND DISCUSSION

4.1 Temporal variation of the NDVI over all land cover categories, 1974-2007.

The temporal variation of the mean NDVI values in March and August was assessed for each land cover category (Fig. 2, Table 2). In both time series there was a clear difference between the vegetated categories (deciduous and coniferous forests, grassland and scrubland) and degraded areas (badlands and erosion risk areas). The vegetated areas had higher NDVI values, and the greatest average NDVI values occurred in August. The NDVI values in March showed positive temporal trends (i.e., the average NDVI increased with time) for all vegetated classes, particularly for deciduous and coniferous forests where the trends were almost significant at the $\alpha = 0.05$ level. Nevertheless, the increase in the NDVI was not constant, and in some years (e.g., 1997 and 2003) a decrease in the average NDVI was detected relative to the general trend (Fig. 2). The NDVI values in August did not show significant temporal trends for any vegetation class. These results suggest an increase in vegetation activity during the study period, especially in March, when the conditions for growth are best.
The degraded areas (badlands and erosion risk areas) had the lowest average NDVI values, which differed little between March and August because of the very low vegetation cover (Table 2). The NDVI trends were negative in both March and August, and were stronger in the erosion risk areas, for which statistical significance was found in the August time series. This may indicate the presence of a degradation process, such as soil erosion, in these areas.
These results suggest that the occurrence of contrasting temporal trends in the overall area depends on the nature of the land cover, with well-vegetated areas undergoing an increase in vegetation activity and degraded areas suffering a process of further degradation. However, the time variability of the NDVI may also be explained by the evolution of climatic conditions, as discussed below.

4.2 Regression analysis of NDVI versus climatic variables

Regression analysis helped explain the observed NDVI temporal patterns of the various land cover classes. The regression models generally fitted the observed NDVI values well, although for pastures, badlands, and erosion risk areas in March, the model results were slightly below the confidence level (Table 3). A better fit was obtained in March for well-vegetated areas (pine and deciduous forests, and scrubland) than for less vegetated regions (pastures, badlands, and erosion risk areas), as shown by the lower R^2 values. In August the goodness-of-fit was similar for all land use classes (Table 6). In all models one or more climatic variables were identified as significant, indicating that climatic conditions were important in explaining the evolution of vegetation activity.

	March				August			
	NDVI		NDVI trend		NDVI		NDVI trend	
Land cover class	mean	sd	rho	p-value	Mean	sd	Rho	p-value
Deciduous forest	0.63	0.10	0.517	0.0862	0.65	0.11	0.321	0.224
Conifers	0.56	0.12	0.573	0.0538	0.61	0.12	0.168	0.520
Grassland	0.49	0.11	0.336	0.281	0.55	0.15	-0.0265	0.926
Scrubland	0.50	0.11	0.294	0.348	0.52	0.13	-0.0899	0.741
Risk erosion areas	0.48	0.11	-0.196	0.543	0.50	0.14	-0.594	0. 0173
Badlands	0.42	0.12	-0.0420	0.904	0.41	0.16	-0.250	0.349

Table 3. NDVI values and temporal NDVI trends (Spearman's rho correlation with time and significance) for each land use category for March and August.

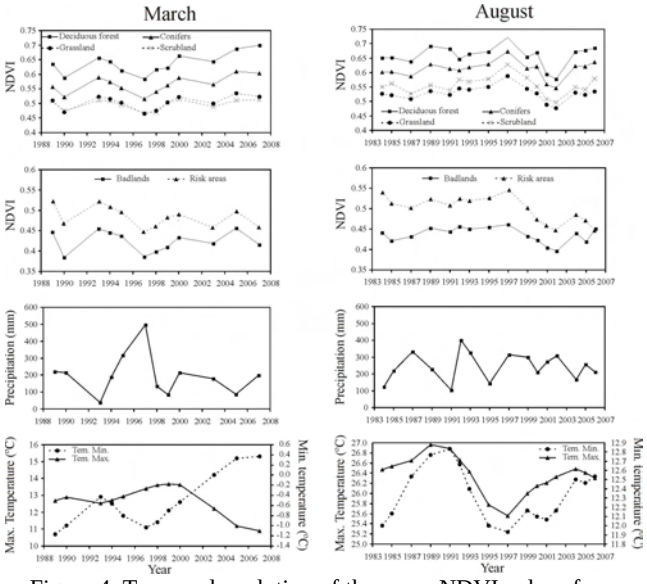


Figure 4. Temporal evolution of the mean NDVI values for March and August between the categories of land cover map and erosion risk areas.

	Pine forest	Deciduous forest	Scrubland	Pastures	Erosion risk	Erosion (badlands)
R^2	0.743	0.779	0.615	0.424	0.467	0.547
p-value	0.002	0.001	0.045	0.084	0.150	0.082
Residual standard error	0.561	0.520	0.728	0.839	0.856	0.789
Beta coefficients:						
Precipitation	-0.317	--	-0.298	--	--	--
T max	--	--	--	--	--	--
T min	0.683	0.678	0.371	1.11	0.701	0.716
Julian day	--	-0.310	-0.326	--	-0.377	-0.457
Time (year)	--	--	--	-0.705	-0.845	-0.719
Temporal trend (change in NDVI):						
per year	--	--	--	-0.00216	-0.00433	-0.00326
period 1989-2007	--	--	--	-4.03%	-7.91%	-6.02%

Table 5. Regression analysis of NDVI values for March in relation to climatic conditions.

	Pine forest	Deciduous forest	Scrubland	Pastures	Erosion risk	Erosion (badlands)
R^2	0.599	0.591	0.663	0.640	0.681	0.663
p-value	0.028	0.031	0.004	0.005	0.003	0.004
Residual standard error	0.739	0.747	0.649	0.671	0.632	0.649
Beta coefficients:						
Precipitation	-0.325	-0.421	--	--	--	--
T max	-1.59	-1.45	-1.66	-1.62	-1.23	-1.50
T min	1.31	1.34	1.21	1.09	0.806	1.42
Julian day	--	--	--	--	--	--
Time (year)	-0.481	-0.507	-0.688	-0.646	-1.000	-0.789
Temporal trend (change in NDVI):						
per year	-0.00197	-0.00247	-0.00241	-0.00244	-0.00363	-0.00210
period 1984-2006	-4.43%	-5.53%	-5.40%	-5.46%	-8.02%	-4.72%

Table 6. Regression analysis of NDVI values for August in relation to climatic conditions.

However, there were differences between March and August, as well as between land cover classes. In March the average minimum temperature was the most important explanatory factor, as evidenced by the fact that this parameter had the largest (standardized) beta coefficient. The effect of the minimum temperature was positive in all cases, with high minimum temperatures yielding elevated NDVI values. This reflects the importance of relatively warm weather at the end of winter/early spring, at the start of the growing period. The maximum temperature was not a significant explanatory factor for any land use class, and cumulative precipitation was significant in only a few cases (pine forest and scrubland). Counter-intuitively, cumulative precipitation had a negative effect on the NDVI (i.e., greater precipitation resulted in lower NDVI), as shown by the negative signs of the beta coefficients. The time of acquisition of the image (variable = 'day') was significant in March for all land cover classes except for pine forest and pasture, suggesting the relevance of the phenological state of vegetation at this time of the year. In August the average minimum temperature was also significant (positive) for all land cover classes, but the most important explanatory factor was the average maximum temperature. This showed the highest absolute beta coefficient and a negative effect in all cases, meaning that a warm summer resulted in lower NDVI values. Cumulative precipitation was significant only for the pine and deciduous forests, where it also had a negative effect on the NDVI. The time of acquisition of the image had no significant effect for any land cover class.

Having thus explained the climatic and phenological effects on NDVI, we proceeded to identify temporal trends in NDVI values for some land cover classes (Tables 3 and 4). Negative time trends were found only in March, for pastures, badlands and erosion risk areas, representing a decrease in the NDVI of 4-8% in the period 1989-2007. Negative time trends were found in August for all land cover classes, and showed a similar range in the period 1984-2006. The magnitudes of the negative trends were similar for all land cover classes with one exception. Erosion risk areas showed the highest values (around 8%) in both March and August.

The results of regression analysis enabled interpretation of the observed temporal patterns in the NDVI (Fig. 2). The apparent upward trend in the NDVI in well-vegetated areas in March can be explained by a similar trend in the average minimum temperature. A downward trend in the NDVI in erosion risk areas was also evident, but was not clearly related to the temporal evolution of any climatic variable.

These results are in agreement with the evolution observed in the western Spanish Pyrenees. Vicente-Serrano et al. (2004) found a general positive trend in the NDVI for forests and well-developed vegetated areas, which was related to an increase in annual mean temperature, and to patterns of land abandonment and natural revegetation processes. In the present study we also found a positive trend in the NDVI for vegetated areas, and showed that the maximum and minimum temperatures in the 3 months before the Landsat images were taken exerted an opposite influence on the NDVI, and that this effect varied during the year.

The finding that cumulative precipitation had a negative effect on the NDVI was puzzling; a positive effect was expected. This anomaly can be explained by the facts that i) water availability is not a limiting factor for vegetation growth in the study area, which receives an average of around 900 mm $year^{-1}$, predominantly in winter and spring; and, ii) the amount of precipitation is well-correlated with cloud cover in the region, with rainy periods resulting in reduced incoming solar radiation, which rises on clear days. It is well known that precipitation level ceases to be a limiting factor for vegetation growth in humid regions, where competition for space and solar radiation is more important. Several studies have documented saturation of the NDVI in relation to precipitation in humid

areas (Santos and Negrín, 1997). In such regions the typically nonlinear relationship between precipitation and vegetation activity could explain the absence of a significant positive effect, but not the presence of a negative effect, as found in our study. However, the second explanation (cloud cover) could assist in an explanation of our data. Unfortunately, neither cloud cover nor terrestrial incoming radiation time series was available, so this hypothesis could not be tested.

In studies of smaller areas, similar NDVI trends in vegetation evolution have been observed, but human impact has been included among the explanations. The presence of a residual temporal trend in NDVI values following removal of climatic influences is usually considered to constitute evidence that other factors, such as human land use practices, are affecting vegetation activity. In the present study we found significant downward trends in all land cover classes in August. Given the low intensity of land use in the region, attribution of these trends to human causes is difficult. Additionally, positive NDVI trends were found in March, which can be explained by the positive evolution of minimum temperatures. Thus, an early start to the growth period, plus increased vegetation activity during spring and early summer, could cause greater stress on vegetation in August. Following exclusion of climatic effects, downward trends in the NDVI were found in March and August for pastures and badland areas, particularly in erosion risk regions. This could be a sign of degradation in such areas because a decrease in the (already low) vegetation activity in the study area has been correlated with soil erosion processes (Alatorre and Beguería, 2009). In erosion risk areas the relative effect of the temporal trend was greater than the effect of any climatic variable, and consequently this land cover type was the only class exhibiting an overall downward trend in the NDVI. As this land cover class includes very sensitive areas that are at risk of loss of all vegetation cover, thus becoming badlands, we focused further on factors that have contributed to this degradation.

4.3 Spatial distribution of positive and negative NDVI trends in erosion risk areas

The downward trend in NDVI for erosion risk areas in March and August could not be explained by climatic factors, and suggested the involvement of degradation processes including active erosion or lateral expansion of existing badlands. This possibility motivated a detailed assessment of the spatial distribution of NDVI trends in erosion risk areas.

Following removal of climatic influences, the spatial distribution of positive and negative trends in the NDVI of erosion risk areas was similar in March and August, indicating that the process is quite consistent and not merely attributable to seasonal effects (Fig. 7). Negative NDVI trends predominated in both images, indicating the occurrence of degradation processes in these areas. However, there were regions in which positive trends dominated, especially in the March image. The proportion of statistically significant trends increased in August, because of an increase in stress conditions, which predominated in this month.

Mapping of trend values on a pixel-by-pixel basis enabled assessment of the importance of particular topographical conditions on the presence of degradation or recovery processes. In March a positive but not statistically significant (p = 0.283) relationship was found between the NDVI trend and elevation (Fig. 8). This may be related to the location of the badland areas; these predominate in the bottom of the Eocene depression, in contrast to forested areas, which are mainly found on slopes. A negative but not significant (p = 0.364) correlation was found between the NDVI trend and the slope gradient, suggesting an association of steeper slopes with more negative trends. This association could be related to the known positive influence of slope gradient on the activity of erosion processes. Similar results were obtained with the August images (Fig. 4), although the relationships with elevation and slope gradient were weaker (p = 0.447 and p = 0.416, respectively).

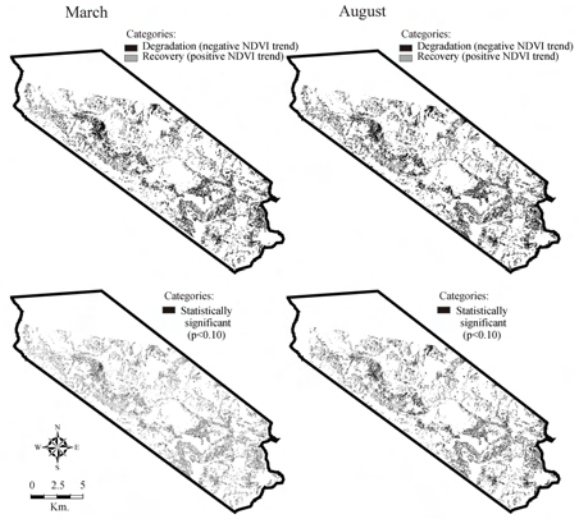


Figure 7. Spatial distribution of the NDVI trends for March and August in erosion risk areas after climatic forcing was accounted for: sign of temporal trend (above) and significance (below).

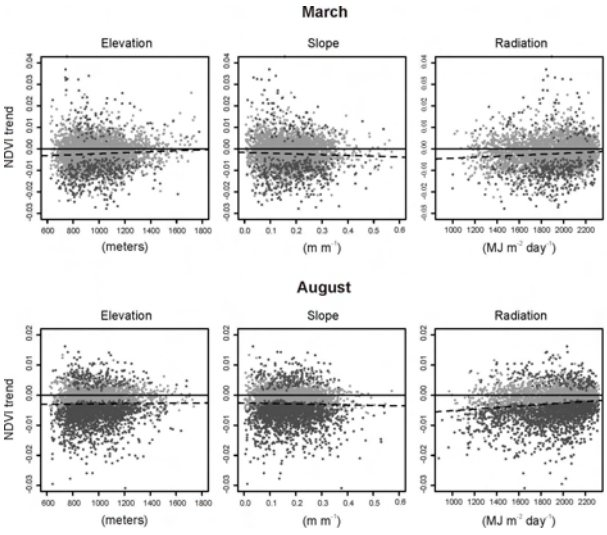


Figure 8. Correlation between the NDVI trends for March and August in the erosion risk areas and topographic variables (after climatic forcing was accounted for). Results are shown for a random sample containing 10% of the original pixels. The black dots indicate pixels with statistically significant trends.

With respect to the potential solar radiation, stronger positive correlations with NDVI trends were found in both the March (p = 0.133) and August (p = 0.0345) image series, suggesting that degradation processes were preferentially occurring on shady (north-facing) slopes. This is consistent with previous research on the topographical signature of badlands in the Spanish Pyrenees, which has revealed that such regions occur predominantly on shady slopes (Alatorre and Beguería, 2009), and are associated with mechanical weathering processes

related to frost and thawing cycles, which are stronger on north-facing slopes (Nadal-Romero et al., 2007). Our results show that the topographic influences on recovery processes are opposite in well-vegetated areas compared to regions undergoing erosion processes.

5. CONCLUSIONS

We analyzed the temporal evolution of vegetation activity on vegetated and degraded surfaces in a small area of the central Spanish Pyrenees over the period 1984-2007. Two map series of the normalized difference vegetation index (NDVI) were obtained from a series of homogenized Landsat TM and Landsat ETM+ images for the months of March and August. This enabled analysis of the spatial and temporal dynamics of vegetation activity in well-vegetated areas (forests and dense scrubland) and degraded areas affected by erosion processes (badlands and risk erosion areas). Temporal NDVI trends were identified for each land cover class using multivariate regression analysis, which incorporated the time evolution of climatic factors (precipitation, and minimum and maximum temperature). Seasonal differences were expected in the spatial pattern of vegetation activity and vegetation recovery processes, as a consequence of the climatic seasonality of the region and the large differences in water availability between spring and summer (vegetation in the latter season is commonly affected by a high level of water stress). The results obtained could have been affected by the heterogeneity of land use and the nature of land covers selected, because this mountainous area is complex and exhibits great spatial diversity. Nevertheless, at the Landsat image spatial resolution (30 m), both land cover and land use were well-represented in the maps. Assignment to class based on the most representative category, by surface area, in a 30 m pixel size could introduce some errors, but it was necessary to guarantee an effective spatial comparison between the NDVI dataset and categorical information. Moreover, the results were spatially consistent, and clear NDVI patterns that coincided with the spatial distribution of land use and land cover were evident. In summary, this study demonstrated that, in a representative mountainous area of the central Spanish Pyrenees, there has been a significant increase in vegetation activity in the last 24 years, which is largely explained by an increase in the minimum temperature. Conifers and deciduous forest have shown the greatest increase in vegetation activity, whereas the increase in activity of grasslands and scrublands has been moderate. Moreover, in active erosion and erosion risk areas, extreme environmental conditions, which accelerate erosion processes, have restricted vegetation recovery processes over this time period.

REFERENCES

Alatorre, L.C., Beguería, S., 2009. Identification of eroded areas using remote sensing in a badlands landscape on marls in the central Spanish Pyrenees. *Catena*, 76, pp. 182-190.

Delbart, N., Picard, G., Le Toan, T., Kergoat, L., Quegan, S., Woodward, I., Dye, D., Fedotova, V., 2008. Spring phenology in boreal Eurasia in a nearly century time-scale. *Global Change Biology*, 14(3), pp. 603-614.

Lambin, E.F., 1996. Change detection at multiple temporal scales: seasonal and annual variations in landscape variables. *Photogrammetric Engineering and Remote Sensing*, 62(8), pp. 931-938.

Lasanta, T., Vicente-Serrano, S., 2007. Cambios en la cubierta vegetal en el Pirieno Aragonés en los últimos 50 años. *Pirineos*, 162, pp. 125-154.

Martínez-Villalta, J., López, B.C., Adell, N., Badiella, L., Ninyerola, M., 2008. Twentieth century increase of Scots pine radial growth in NE Spain shows strong climate interactions. *Global change biology*, 14(12), pp. 2868-2881.

Nadal-Romero, E., Regüés, D., Martí-Bono, C., Serrano-Muela, P., 2007. Badlands dynamics in the Central Pyrenees: temporal and spatial patterns of weathering processes. *Earth Surfaces Processes and Landforms*, 32(6), pp. 888-904.

Pons, X., Ninyerola, M., 2008. Mapping a topographic global solar radiation model implemented in a GIS and refined with ground data. *International Journal of Climatology*, 28, pp. 1821-1834.

Pueyo, Y., Beguería, S., 2007. Modelling the rate of secondary succession after farmland abandonment in a Mediterranean mountain area. *Landscape and Urban Planning*, 83(4), pp. 245-254.

Riaño, D., Ruiz, J.A.M., Isidoro, D., Ustin, S.L., 2007. Global spatial patterns and temporal trends of burned area between 1981 and 2000 using NOAA-NASA Pathfinder. *Global Change Biology*, 13, pp. 40-50.

Rouse, J.W., Hass, R.H., Schell, J.A., Deering, D.W., Harlan, J.C., 1974. Monitoring the vernal advancement and retrogradation (greenwave effect) of natural vegetation. NASA/GSFC type III final report, Greenbelt, M.D.

Santos, P., Negrín, A.J., 1997. A comparison of the Normalized Difference Vegetation Index and rainfall for the Amazon and Notheastern Brazil. *Journal of Climate*, 36, pp. 958-965.

Vicente-Serrano, S.M., Beguería, S., López-Moreno, J., García-Vera, M., Stepanek, P., 2009. A complete daily precipitation database for North-East Spain: reconstruction, quality control and homogeneity. *International Journal of Climatology*, DOI: 10.1002/joc.1850

Vicente-Serrano, S.M., Lasanta, T., Romo, A., 2004. Analysis of the spatial and temporal evolution of vegetation cover in the Spanish central Pyrenees: the role of human management. *Environ. Manage*, 34, pp. 802-818.

Vicente-Serrano, S.M., Pérez-Cabello, F., Lasanta, T., 2008. Assessment of radiometric correction techniques in analyzing vegetation variability and change using time series of Landsat images. *Remote Sensing of Environment*, 112, pp. 3916-3934.

Warren, A., 2002. Land degradation is contextual. *Land Degradation and Development*, 13, pp. 449-459.

ACKNOWLEDGEMENTS

This research was financially supported by the project CGL2006-11619/HID, funded by CICYT, Spanish Ministry of Education and Science.

ESTIMATING SUB-PIXEL TO REGIONAL WINTER CROP AREAS USING NEURAL NETS

C. Atzberger [a,*], F. Rembold [a]

[a] Joint Research Centre of the European Commission, IPSC, Via Enrico Fermi 2749, 21027 Ispra (VA), Italy – (clement.atzberger, felix.rembold)@jrc.ec.europa.eu

KEY WORDS: Neural, Pixel, Estimation, Method, Test

ABSTRACT:

The work aimed at testing a methodology which can be applied to low spatial resolution satellite data to assess inter-annual crop area variations on sub-pixel to regional scales. The methodology is based on the assumption that within mixed pixels land cover variations are reflected by changes in the related hyper-temporal profiles of the Normalised Difference Vegetation Index (NDVI). We evaluated if changes in the fractional winter crop coverage are reflected in changing shapes of annual NDVI profiles and can be detected by using neural networks. The neural nets were trained on reference data obtained from high resolution Landsat TM/ETM images. The proposed methodology was applied in a study region in central Italy to estimate winter crop areas between 1988 and 2002 from 1 km resolution NOAA-AVHRR profiles and additional ancillary data readily available (CORINE land cover). The accuracy of the estimates was assessed by comparison to official agricultural statistics using a bootstrap approach. The method showed promise for estimating crop area variation on sub-pixel level (cross-validated R^2 between 0.7 and 0.8) to regional scales (normalized RMSE: 10%). The network based approach proved to have a significantly higher forecast capability than other methods used previously for the same study area.

1. INTRODUCTION

There is a growing concern for large-scale environmental issues such as global warming, loss of biodiversity and food security. Remote sensing is the only practical source of environmental data with global coverage. Moderate to coarse spatial resolution satellite sensors such as NOAA-AVHRR or SPOT VGT provide synoptic information at a high temporal resolution while the amount of data is still manageable. Due to their coarse resolution, however, most pixels contain a mixture of land cover classes, referred to as sub-pixel mixing (Atkinson et al., 1997).

Knowledge of the spatial distribution of crop types is important for land management and trade decisions, and is needed to regularize the inversion of radiative transfer models for mapping crop biophysical and biochemical variables (Atzberger, 2004; Richter et al. 2009). However, in regional to global agricultural studies, mixed spectral signatures are common (Lobell & Asner, 2004). For example, small field sizes (1 - 10 ha) are typical for many agricultural systems in the developing world, but also in Europe. Thus, methods are required that allow the mapping of fractional coverages from coarse resolution imagery (Foody et al., 1997). If suitable methods could be developed, the analysis of archived images would also allow tracing back the development of the landscapes to the early eighties, when the first global data sets became available.

To un-mix coarse resolution imagery, many studies rely on the assumption of a linear relationship between end-member signatures and the composite signature (e.g. Quarmby, 1992). However, mixing is often non-linear and end-member spectra are sometimes difficult to obtain.

To overcome these problems, Foody et al. (1997) proposed an un-mixing approach based on neural nets. The approach makes no assumptions about the nature of the mixing and does not require end-member spectra. Relative to a conventional classification oriented approach, the areal extent of classes was generally more accurately estimated from (single date) AVHRR data after the application of the unmixing procedure.

Neural networks (NN) were also used together with mixture modelling and fuzzy classification by Atkinson et al. (1997) for mapping sub-pixel proportions of land cover classes in U.K. Again, it was found that neural networking was the most accurate technique, but its successful implementation depended on accurate co-registration of the 5 band AVHRR image with a high resolution SPOT HRV image. The availability of an accurate training data set was also very important.

A probabilistic temporal unmixing (PBU) using MODIS reflectance data was proposed by Asner & Lobell (2004). Landsat data was used to identify pure pixels for the extraction of temporal endmembers. Sub-pixel fractions of crop area were modeled by using linear un-mixing. Performance of the mixture model varied from 50 to 80 % depending on the scale of comparison.

NDVI time series from NOAA-AVHRR was used successfully by Remold & Maselli (2004; 2006) for estimating inter-annual crop area changes in Tuscany (Italy). The study used an approach based on the Spectral Angular Mapping (SAM). For model calibration, test data derived from high resolution Landsat TM and ETM was used. Albeit relatively accurate inter-annual crop area changes could be achieved, the results were strongly dependent on the quality of the satellite images. Useful results for winter crops

* Corresponding author

Figure 1. Study area Tuscany, Italy. The sample image is from AVHRR (top left: 44° 42'-8° 19', lower right: 42° 00'-13° 21')

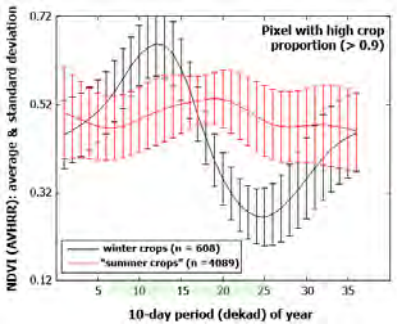


Figure 2. NDVI profiles of pure winter and "summer crops". The dekads used for modelling are shown in green

could be provided only by end of September, reducing considerably the timeliness of the area change information.

The use of neural nets for estimating sub-pixel land cover from temporal signatures was investigated by Verbeiren et al. (2008). Monthly MVC of SPOT-VGT (between March and October) were used to model the area fraction images (AFI) of eight classes in 2003 for Belgium. Relatively good results were obtained especially if the initial (pixel-based) results were aggregated to higher regional levels. The portability of the trained networks across growing seasons was investigated by Bossyns et al. (2007) in an accompanying paper on the same data set. The NNs were trained on data of one growing season and than applied to SPOT-VGT of the training year and three other seasons. High and stable accuracies of the estimated AFI's were obtained when the trained network was applied on the imagery of the training year. For example, at regional level, the R^2 for winter wheat was ~0.8 (0.67-0.87) for the training years. However, on average, this values decreased by 0.45 units when the trained networks were applied to different seasons, probably because of a too high inter-annual variability of the temporal signatures.

The objectives of the present study were:

- to test if NNs can be used with low resolution NOAA-AVHRR imagery and additional ancillary information to accurately estimate winter crop surfaces at sub-pixel to regional scales between 1988 and 2002
- to evaluate the impact of ancillary land use information (CORINE) on the estimation accuracy as well as the influence of an improved smoothing of the AVHRR time series
- to determine the optimum prediction dates for early-season forecasts of winter crop surfaces

2. MATERIAL

The methodology was tested using data for the Tuscany region in Central Italy. The choice was driven mainly by the availability of both satellite imagery and agricultural statistics. The region is covered by a consistent NOAA-AVHRR data time series taken in the period from 1986 to 2003, when also several Landsat TM/ETM+ scenes were acquired. An area frame sampling method has been regularly applied since 1988 to measure the extent of the main crops in Tuscany (Carfagna et al. 1998).

2.1 Geography and environmental features of the study area

Tuscany is situated between 9°- 12° East longitude and 44°- 42° North latitude, covering circa 2 x 10^6 hectares (figure 1). From an environmental point of view, Tuscany is peculiar for its extremely heterogeneous morphological and climatic features. The topography ranges from flat areas near the coast-line and along the principal river valleys to hilly and mountainous zones towards the Apennine chain. The climate in Tuscany ranges from typically Mediterranean to temperate warm or cool according to the altitudinal and latitudinal gradients and the distance from the sea.

The land use of Tuscany is predominantly agricultural where the land is flat and mixed agricultural and forestry in the hilly and mountainous areas. The main agricultural cover types are cereal crops in the plains and olive groves and vineyards on the hills. The upper mountain zones are almost completely covered by pastures and forests. Cropland is spread over the coastal zones and the inner plain and hilly areas cover approximately 25% of the land surface. The prevalent cereal is durum wheat, with an average planted area of 112 000 ha and with a mean growing cycle from November to the end of June (figure 2).

2.2 Data

2.2.1 Reference information on winter crop areas: The land cover classification of Tuscany produced by the CORINE project was used as reference map (Annoni & Perdigao, 1997). Wheat area estimates for the period 1988-2002 were obtained from the AGRIT project (Consorzio ITA, 1987). These statistics are produced annually through an area frame sampling method, which guarantees high estimation accuracy (error < 10 %) at the regional scale (Carfagna et al., 1998). From the available data, 1994 has been excluded because of the insufficient quality of the AVHRR data. In what follows, we use the term "winter crop area" as a synonym for the wheat area.

2.2.2 High resolution images: The high resolution images were necessary to spatialize the statistical information provided by the AGRIT statistics (2.2.1). The high resolution data set consisted of 8 Landsat frames (192/30), 6 taken by the Thematic Mapper (TM) (1988, 1991, 1992, 1995, 1997 and 1998) and 2 by the Enhanced Thematic Mapper (ETM+) (2000, 2001). All of them were acquired during the month of August and were cloud free over the main agricultural areas. The Landsat TM/ETM+ scenes were first geo-referenced by a nearest neighbor resampling algorithm using

	Total winter crop surface (Tuscany)		Sub-pixel winter crop surfaces, i.e. spatialization	
Year	AGRIT [ha]	NNcv [ha]	RMSEcv (RMSE)	R^2cv (R^2)
1988	133795	147543	9,6 (8,8)	0,70 (0,80)
1991	184024	152491	10,8 (8,8)	0,79 (0,78)
1992	112450	96053	8,3 (9,0)	0,71 (0,79)
1995	145999	170981	8,8 (8,9)	0,81 (0,79)
1997	167000	136426	10,3 (8,8)	0,72 (0,80)
1998	174296	168462	9,0 (8,8)	0,79 (0,79)
2000	183356	183569	9,8 (8,9)	0,82 (0,77)
2001	154510	201407	10,6 (8,9)	0,77 (0,79)
X_{med}	160755	160477	9,4 (8,9)	0,79 (0,79)

Table 3. Cross-validated results obtained with the neural network for the eight years for which reference information was available. The results for the training data are given in parentheses

about 120 control points selected on a CORINE-based land/water mask. Bands 4 (nIR) and 3 (Red) were corrected for atmospheric effects and converted into reflectances from which high resolution NDVI images for every training year were calculated.

2.2.3 Low resolution images: JRC-MARS owns the most elaborate archive of NOAA-AVHRR 1km data over the pan-European continent. In 2008, all historical data were re-processed with new procedures, which resulted in an unique archive of 27 years. For this study, the AVHRR time series from 1988 to 2002 was used, fully encompassing the years for which high resolution TM/ETM data was available (2.2.2).

3. METHODS

3.1 Generation of reference abundance maps

In a first step the CORINE land cover classes were grouped into five environmentally meaningful categories (summing up to unity) with more or less similar NDVI profiles. Besides the "arable land" class, four other categories were derived: forests, pastures, tree plantations and urban areas (Maselli, 2001). We assumed that the broad land cover was stable between 1988 and 2002. In a second step, the CORINE "arable land" class was split into winter crops and summer crops using the available TM/ETM+ images of the eight training years (2.2.2). For this purpose, each high resolution NDVI image was thresholded. The threshold was determined so that the summed high resolution winter crop surfaces equal the AGRIT statistic (2.2.1). This operation was possible as winter crop fields are almost bare in August, while other fields with summer crops (maize, sunflower, etc.) but also fallows and pastoral areas are in a "green" phase. This also implies that the class "summer crop" is a mixture of typical summer crops with other classes. Although the thresholding resulted in two masks per image (one for the winter crops, the other showing the distribution of summer crops), the latter was not further used.

The five categories identified in the eight training dates (four CORINE categories plus winter crops) were spatially degraded by pixel averaging to produce abundance images (i.e., AFI) with the same spatial resolution as the AVHRR images. Of course, out of these images only the winter crop abundance maps were different for each date. Hence, only the variations within the CORINE "arable land" category were analyzed, without considering any changes within the other categories.

	Total winter crop surface for Tuscany		
	Nobs	R^2	nRMSE
training years	8	0.77	7.9
validation years	6	0.35	12.4
pooled data	14	0.57	10.7

Table 4. Results obtained with the neural network trained on the eight years for which reference information was available

3.2 Smoothing of low resolution AVHRR data

Time series from AVHRR require a careful filtering/smoothing before they can be applied within quantitative studies (Beck et al., 2006). The standard maximum value compositing (MVC) only corrects for major disturbances. To eliminate the negatively biased noise typical for coarse resolution time series, the modified Whittaker smoother was used (Atzberger & Eilers, 2010a; 2010b). The filter was chosen because it is very fast, interpolates easily and optimizes its smoothness parameter automatically.

3.3 Neural networking

A simple net with one hidden layer was used to map the winter crop fraction from the profiles of coarse resolution NDVI images and the ancillary data (e.g. five AFI related to arable land, forests, tree plantations, urban areas and pastures). The output layer represented the winter crop area fraction and had thus only one neuron. In the standard setting, the number of input neurons was 21. This allowed to simultaneously input the 5 (inter-annually constant) abundance values derived from the CORINE data plus 16 neurons describing the temporal NDVI profile (dekads 7 to 25, excluding dekads 14 to 16). The three dekads were excluded because winter and summer crop signatures strongly overlap during this period (figure 2). The number of neurons in the hidden layer was set to 3, resulting in a compact 21-3-1 network architecture. For network training the resilient backpropagation algorithm was used. To improve generalization of the net and to prevent overfitting, the early stopping technique was applied (Atkinson & Tatnall, 1997; Farifteh et al. 2007). For this purpose, the training samples were split into three subsets with 50 % (training), 25 % (test) and 25 % (validation) of the total available pattern. Only the training set was used for computing the gradient and up-dating the weights and biases. The training was stopped automatically when the test error started to rise and the actual weights were returned. To keep the methodology simple, other more elaborated training strategies were not investigated.

4. RESULTS AND DISCUSSION

4.1 Spatio-temporal distribution of winter crops

To derive statistically sound results, a jacknife procedure was selected. This means that 8 different training sessions were run, each time using only 7 out of the available 8 years. The left-out sample was than predicted by the trained network. The resulting

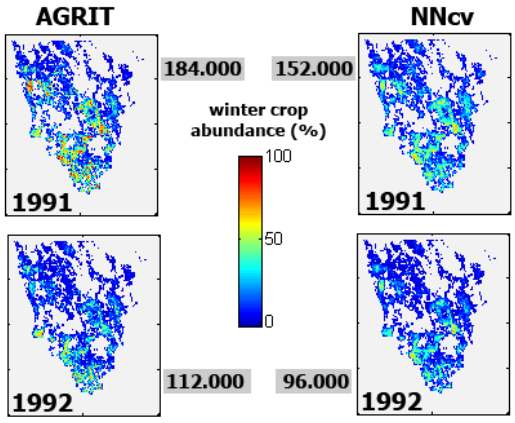


Figure 5. Reference (left) and cross-validated NN-derived (right) spatial distribution of winter crop surfaces for 2 contrasting years

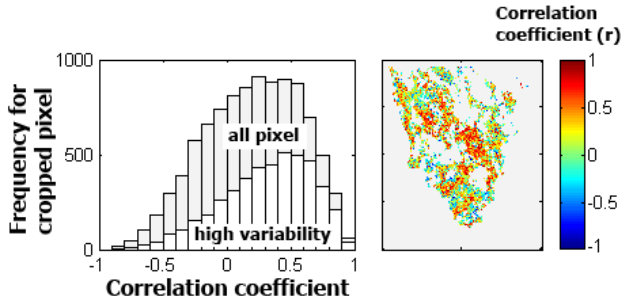


Figure 6. Spatial distribution of the cross-validated correlation coefficient (r) between AGRIT reference information and the modeled results (right). The corresponding frequency distribution is shown on the left. The histogram gives the frequency for all pixels (in gray) and for those having a high inter-annual winter crop area variability (in white)

winter crop surfaces for Tuscany are reported in table 3. The table also lists two statistics describing the accuracy with which the (spatial distribution) of the abundance images was modeled, distinguishing between cross-validated results and training data.

On average, 79 percent of the observed spatial variability of the (sub-pixel) winter crop area was explained by the net. However, we also observe large variations from year to year with R^2cv ranging between 0.70 (1988) and 0.82 (2000). In general, the cross-validated RMSE values of winter crop areas were around 10 %. As expected, cross-validated RMSE were higher than those obtained on the training sample (< 9 %). Together, this indicates that the spatial distribution of winter crops was quite well modeled, however, with sometimes some significant offset.

Spatialized winter crop abundances are shown in figure 5 for the two most contrasting years of the data set (1991 and 1992). A high winter crop surface was reported for 1991. In 1992, the winter crop surface dropped by almost 40 % as a reaction of new European policies. The NN well depicted this general tendency. The spatial distribution of winter crop surfaces was also well modeled (figure 5) with, however, a noticeable bias.

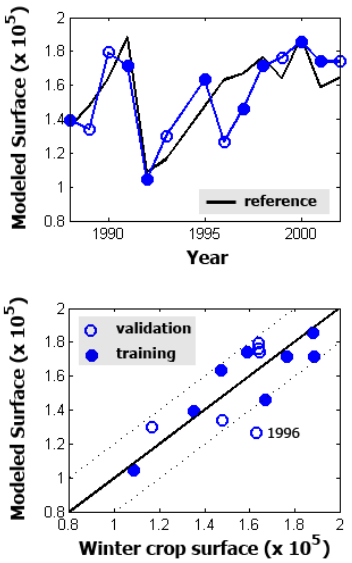


Figure 7. Total winter crop area (1988-2002; 1994 excluded) for Tuscany modeled by the NN against reference data (AGRIT). The eight training years (filled circles) were used for network training

The cross-validated correlation coefficient (r) between reference winter crop fraction (AGRIT) and the modeled value is shown in figure 6 (right) for each AVHRR pixel for which CORINE indicated at least 2 percent arable land. Two frequency distributions are shown as well (left): (i) for all pixels with arable land, and (ii) excluding those pixels having a low inter-annual variability in winter crop area (STD < 5%). The figure shows that the NN often performed well to estimate the inter-annual variability of winter crop areas at this high dis-aggregation level. The amount of inter-annual variance explained by the net increased when stable pixels were excluded. However, low (or even negative) correlations were frequently obtained in the plains of Tuscany (blue colors). The net probably failed because the "summer crop" signatures were too variable from year to year in these areas.

We tested if an increased number of neurons in the hidden layer could improve the results. This was not the case (Atzberger & Rembold, 2009). Albeit it was possible to increase the accuracy within the training set, the net could not generalize this improvement for yielding better predictions on the left-out samples.

4.2 Regional winter crop area estimates 1988-2002

To evaluate the capacity of the net to predict regional winter crop surfaces of Tuscany for the entire AVHRR time series, the NN was trained with all 8 years for which reference information was available. The trained net was next applied to the full time series (1988 to 2002). The results are shown in table 4 and figure 7.

Overall, the regional winter crop area was well modeled with a normalized RMSE of around 10 percent. Noticeable, however, is the strong decrease in performance for the six years not included in the training set. For this data, only 35 percent of the inter-

	Assessment of CORINE and filtering impact		
	Dataset 1 ("standard")	Dataset 2 ("no ancillary")	Dataset 3 ("no filtering)
CORINE	Yes	No	No
Filtering	Yes	Yes	No
Inputs	21	16	16
R^2	0.60	0.52	0.40
nRMSE	10.0	14.4	14.1
Nobs	14	14	14

Table 8. Statistics obtained on the pooled data set (14 years of data – 1994 excluded) from which 8 years where used for network training. The three data sets differ in the amount of ancillary information (CORINE land cover) and the pre-processing applied to the NOAA-AVHRR data. The statistics refer to the accuracy with which the regional winter crop surfaces were modeled

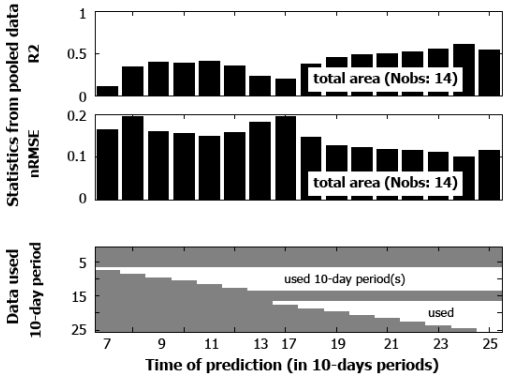


Figure 9. Accuracy of the net to predict regional winter crop areas (pooled data set; Nobs: 14). The two bar charts display the statistical information. The third chart indicates which images were used (left: early predictions; right: late predictions)

annual variance was explained by the NN (compared to 79 percent of the training years), mainly because 1996 was seriously underestimated. This year was also not well modeled in the precursor study using SAM (Rembold & Maselli, 2006). Until now we were unable to indentify the exact reasons that led to this outlier.

4.3 Impact of CORINE data and the smoothing of AVHRR data

One of the study objectives was to evaluate the impact of the ancillary data (CORINE) and the filtering of the AVHRR data. Table 8 reports the main results referring to the regional winter crop surfaces of the 14 years 1988-2002. The (positive) impact of the CORINE data was strong. The R^2 decreases from 0.60 to 0.52 when the five CORINE abundance maps were not used. At the same time the normalized RMSE increased from 10 to more than 14 %. The R^2 value further decreased if the original (unfiltered) NDVI data were used instead of the filtered images.

4.4 Within-season predictions

Early (within-season) predictions of crop surfaces are of upmost importance in any agricultural monitoring system such as MARS (Lobell & Asner, 2004). To evaluate the possibility to estimate the winter crop area of Tuscany already early in the season, 16 different nets were trained on the 8 reference years and applied to the pooled data set (14 years). The nets solely differed by the number of low resolution NDVI images used as network input. Figure 9 summarizes the results, where the plot on the bottom specifies the used (and excluded) NDVI data. The two bar charts indicate the statistics obtained more or less early in the growing season. Not surprisingly, the best results were obtained if NDVI data from the full winter crop cycle (e.g., up to dekad 25) was used for the modelling ($R^2 > 0.5$, nRMSE < 11%). But already by March-April some relatively accurate first predictions of winter crop (and potential summer crop) areas were achieved. The accuracy was substantially lower compared to the predictions made at the end of the winter growing season, but could possibly be useful. It has also to be noted that the analyzed data set still contained the data from 2006 which was previously identified as problematic (4.2).

5. CONCLUSIONS

The study evaluated the performance of neural networks to map (1) the spatial distribution of winter crops, and (2) the inter-annual variation in regional winter crop surfaces. The nets were fed by low resolution (NOAA-AVHRR) imagery and ancillary information (CORINE). Reasonable results were obtained with a compact standard backpropagation network (21 inputs, 3 hidden neurons). The results were better than those obtained in a previous study with the same data set using the SAM of temporal NDVI profiles (Rembold & Maselli, 2006). Albeit the performance of the net decreased when applied to different growing seasons, the drop in accuracy was not as strong as in Bossyns et al. (2007). Hence, the network generalized comparatively well. Considering the spatialization of the winter crops, it has to be noted that the accuracy of the reference (TM/ETM) images is unknown. Errors in these reference maps will automatically deteriorate the training process and fudge the accuracy assessment.

Using the same data set, the neural net gave relatively good results much earlier than in our previous work (Rembold & Maselli, 2006). In the precursor study reasonable results were only obtained after August, whereas in the present study first reasonable results could already been achieved end of March, making the proposed method more interesting for in-season crop monitoring and area estimation.

The research clearly demonstrated the positive impact of using ancillary information (CORINE) in the modelling. With the help of this information (albeit known not to be perfect), the neural net was able to better 'learn' the relation between the temporal signatures and the corresponding area fraction images (AFI) of the winter crops. This result looks plausible, as the proportion and composition of non-arable classes within a pixel can vary significantly. This will inevitably affect the resulting NDVI profile. By providing the NN with ancillary information, more specific mapping functions can be developed.

The main limitation of the proposed approach relates to the variability of signatures which are not winter crops (e.g., the "summer

crops"). Indeed, an analysis of pixels with >90% arable land and labeled as "summer crop" revealed signatures that were highly variable in space and time (not shown). The consequences were clearly seen in several plains of Tuscany where it is known that the proportion, composition and phenology of summer crops have a high inter-annual and spatial variability. To tackle this problem, a follow-up study will try to "normalize" the NDVI profiles by taking meteorological indicators and DTM information into account. Likewise, one could try to improve network generalization by providing the net with an additional output variable indicating the area fraction of summer crops.

1996 was the year presenting the highest problems for winter crop area estimation (figure 7). Further investigation is needed to understand exactly what makes 1996 different from the other years. The additional research should address both the quality of the AGRIT statistics for this particular year and possible climatic or agronomic factors. It is curious however to observe that the same year was also the worst estimate in the precursor study, indicating that independently from the methodology used, it is difficult to explain Tuscany's winter crop area of 1996 by using low resolution NDVI data.

Considering the promising results obtained, it would be of interest to test the proposed methodology for its robustness in other geographical areas and larger regions as well as by using other low or medium resolution NDVI time series such as SPOT VGT or MODIS. The main requirements for further investigations in this sense are the availability of:

- training data for the discrimination between winter and summer crops, either based on high resolution data analysis or on existing agricultural databases (e.g. high resolution land use classifications or cadastral data; Verbeiren et al., 2008);
- reliable crop area statistics for model validation, possibly based on a high accuracy area frame sampling approach, such as the one of the AGRIT project (Consorzio ITA, 1987) at provincial or national level.

REFERENCES

Annoni, A., Perdigao, V., 1997. *Technical and methodological Guide for updating CORINE Land Cover Data Base*. European Commission, EUR 17288EN, Ispra, Italy, 124 p.

Atkinson, P.M., Cutler, M.E.J., Lewis, H., 1997. Mapping sub-pixel proportional land cover with AVHRR imagery. *Int. J. Remote Sens.*, 18(4), pp. 917-935.

Atkinson, P.M., Tatnall, A.R.L., 1997. Neural networks in remote sensing. *Int. J. Remote Sens.*, 18(4), pp. 699-709.

Atzberger, C., Eilers, P.H.C., 2010a. Evaluating the effectiveness of smoothing algorithms in the absence of ground reference measurements. *Int. J. Remote Sens., accepted*.

Atzberger, C., Eilers, P.H.C., 2010b. A smoothed 1-km resolution NDVI time series (1998-2008) for vegetation studies in South America. *under review , Int. J. Digital Earth*.

Atzberger, C., Rembold, F., 2009. Estimation of inter-annual winter crop area variation and spatial distribution with low resolution NDVI data by using neural networks trained on high resolution images. In: *Proc. SPIE Remote Sensing*, Berlin, Germany, DOI: 10.1117/12.830007

Beck, P., Atzberger, C., Høgda, K., Johansen, B., Skidmore, A., 2006. Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI. *Remote Sens. Environ.*, 100, pp. 321−334.

Bossyns, B., Eerens, H., van Orshoven, J., 2007. Crop area assessment using sub-pixel classification with a neural network trained for a reference year. In: *Proc. 4th intern. workshop on the analysis of multi-temporal remote sensing imagery*, Belgium, pp. 1-8.

Carfagna, E., Ragni, P., Balli, F., 1998. Crop production agricultural survey based on area frame sampling methods (1988-1997). In: *FAO Statistical Development Series*, 10(13), pp. 203-216

Consorzio ITA, 1987. *Telerilevamento in Agricoltora, Previsione delle Produzioni di Frumento in Tempo Reale e Sviluppi Tecnologici*. Ministero dell'Agricultura, Roma, Italy.

Farifteh, F., Van der Meer, F., Atzberger, C., Carranza, E.J.M., 2007. Quantitative analysis of salt-affected soil reflectance spectra: A comparison of two adaptive methods (PLSR and ANN). *Remote Sens. Environ.*, 110, pp. 59-78.

Foody, G.M., Lucas, R.M., Curran, P.H., Honzak, M., 1997. Non-linear mixture modeling without end-members using an artificical neural net. *Int. J. Remote Sens.*, 18(4), pp. 937-953.

Lobell, D.B., Asner, G.P., 2004. Cropland distributions from temporal unmixing of MODIS data. *Rem. Sens. Env.*, 93, pp. 412-422.

Maselli, F., 2001. Definition of spatially variable spectral endmembers by locally calibrated multivariate regression analysis. *Remote Sens. Environ.*, 75, pp. 29-38.

Quarmby, N.A., 1992. Linear mixture modelling applied to AVHRR data for crop area estimation. *Int. J. Remote Sens.*, 13(3), pp. 415-425.

Rembold, F., Maselli, F., 2006. Estimation of inter-annual crop area variation by the application of spectral angle mapping to low resolution multitemporal NDVI images. *Photogrammetric Engineering and Remote Sensing*, 72(1), pp. 55-62.

Richter, K., Atzberger, C., Vuolo, F., Weihs, P. and D'Urso, G., 2009. Experimental assessment of the Sentinel-2 band setting for RTM-based LAI retrieval of sugar beet and maize. *Can. J. Remote Sens.*, 35(3), pp. 230-247.

Verbeiren, S., Eerens, H., Picard, I., Bauwens, I., van Orshoven, J., 2008. Sub-pixel classification of SPOT-VEGETATION time series for the assessment of regional crop areas in Belgium. *Int. J. Appl. Earth Observation*, 10, pp. 486-497.

INTEGRATION OF CONTEXTUAL INFORMATION FOR THE TRANSFER OF BELIEFS IN AN INFORMATION SOURCES FUSION SYSTEM – APPLICATION TO DETECTION AND CLASSIFICATION OF TREES CROWNS

Z. Ben Dhiaf [a,] *, A. Hamouda [a], J. Desachy [b]

[a] Faculty of Science of Tunis, Unit of Research in Programming, Algorithmic and Heuristic – {zouhour_bendhiaf, atef_hammouda}@yahoo.fr
[b] University of the Antilles and Guyana, Groups of Research in Data processing and Mathematics Applied - jdesachy@univ-ag.fr

KEY WORDS: Information, Fusion, Image, Classification, Forestry

ABSTRACT:

In this paper, we present an approach based on the transferable belief model for the detection and the classification of trees crowns on very high resolution satellite images of forest scenes. The masses resulting from the high resolution image source don't always allow deciding between the classes of occupation satisfactorily. Forest context and information about the structure of the forest species are two key elements in the forest scenes classification process. We expose in this paper the retained modelling of the context concept and the approach of revising masses through a transfer of belief. Then, we give some experimentation that illustrates the given approach.

1. INTRODUCTION

Remote sensing images classification is confronted to the theoretical requirements of information sources fusion approaches. The existing models offer an interesting theoretical framework (J. Desachy and all, 2000). However, their application often emphasizes the incapacity of these methods to take into account the context. Context is traduced by determinant expert's knowledge and is considered in this paper not as elementary information but rather as a contextual constraint in the masses calculations specially those corresponding to belief masses functions revision. Indeed, analysis of the image pixel by pixel, often adopted as the basis of the classification process does not take into account the object to which the pixel belongs. But this information, that we call "contextual information", may be important for the classification process. It is the case of forest scenes images where the trees crowns delimit a subset of classes of occupation. The mass estimation of a scene point will thus have to take into account if this point is inside or outside the crown.
This paper exposes an approach allowing taking into account the context in the evaluation of the masses. We present also the manner with which we modeled this problem. An application on high resolution images (HRI) emphasized results very close to the field reality.
The approach that we propose uses belief functions theory as fusion formalism. We apply a beliefs transfer based on contextual information. In the following section, we tried to summarize the basic notions of both belief functions theory and transferable belief model. These concepts are essential for the explanation of the stages of our approach detailed in section 3.

2. BELIEF FUNCTIONS THEORY

The belief functions theory was named at the beginning with the name of its authors: Dempster and Shafer (Shafer, 1990).

The origin of belief functions theory started with the works of Arthur P. Dempster. Those works are related to the statistical inference theory generalizing the Bayesian inference. G. Shafer proposed belief functions as general framework of representation of uncertainties, including the probabilities theory like particular case. Extensions to the Dempster–Shafer theory (DST) contributed to the enrichment of the belief functions theory (Bloch 2005; Bloch 1996; Denoeux 2004).

Ph. Smets suggested a model named transferable belief model (TBM) providing coherent non-probabilistic interpretation of the DST and clarifying the concept subjacent with it (Smets, 1990).

The belief functions theory is one of the theories largely used for information sources fusion considering the fact that it takes into account simultaneously sources uncertainties and provided information inaccuracy. It is reduced to the theory of probability and the theory of the possibilities in particular cases (Burrus 2003, Vannoorenberghe, 2003).

2.1 Information sources and power set

Each source of information being in general imperfect, it is significant to combine several sources in order to have better knowledge of the "world". We will consider in the continuation that we have *n* sources of information S_i with $i \in \{1,\ldots,n\}$.

Those sources must make a decision on an observation x in a whole of k decisions $C_1,\cdots,C_k$. Let $\Omega = \{C_1,\cdots,C_k\}$ being the frame of discernment composed of k hypotheses (exclusive and exhaustive), 2^Ω is the power set (it is the set of parts of Ω ($2^\Omega = \{A_i | A_i \subseteq \Omega\}$) and the A_i are the events of 2^Ω with $i \in \{1,\ldots,|2^\Omega|\}$.

* Corresponding author

2.2 Belief mass functions

The belief mass function $m(A)$ (or simply mass function) of an event A, is the confidence carried strictly in A without being able to be divided onto the hypotheses composing A. The focal elements are the elements of 2^{Ω} of non null masses. If the source is perfect, information is precise and sure, there is thus a single hypothesis C_i such as $m(\{C_i\})=1$.
The mass functions are then defined on each subset of the set of disjunctions of 2^{Ω} to values in [0,1]. The distribution of mass is written according to (1) :

$$\begin{aligned} m: 2^{\Omega} &\rightarrow [0,1] \\ A &\rightarrow m(A) \end{aligned} \qquad (1)$$

Dempster proposes a conjunctive rule of combination between sources called orthogonal sum. This combination causes to assign the masses to propositions of which the number of elements is less than that of the original propositions. For two sources S_1 and S_2, one writes the orthogonal sum $\oplus$, in the following form (2):

$m = m^{S_1} \oplus m^{S_2}$, which is written for an event A like:

$$m(A) = \sum_{B \cap C = A} m^{S_1}(B) \times m^{S_2}(C) \qquad (2)$$

Evidential modelling makes it possible to represent at the same time the inaccuracy and uncertainty through two functions of credibility and plausibility, derived from the mass functions. The decision is done by maximization of one or the other of these two functions.

2.3 Transferable Belief Model

In this model, two levels can be distinguished: the credal level where the beliefs are modelled and revised, and the pignistic level in which the belief functions are transformed into probability functions, known as pignistic ($BetP$), for the decision-making (Smets, 1990).
Maximum of pignistic probability is generally considered with singletons hypothesis C_j because of the additivity of probabilities. With singleton hypotheses we obtain the following equation (3):

$$BetP(C_j) = \frac{1}{1-K} \sum_{C_j \in A_i, A_i \neq \phi} \frac{m(A_i)}{|A_i|} \qquad (3)$$

With K is the conflict between sources expressed usually as the mass of the empty set ϕ like illustrated equation (4):

$$K = m(\phi) = \sum_{\substack{A \cap B = \phi \\ A \in 2^{\Omega}, B \in 2^{\Omega}}} m^{S_1}(A) \times m^{S_2}(B) \qquad (4)$$

The name "Transferable Belief Model" (TBM) comes from the transfer of belief allocated initially in a proposition towards a more specific subset of it. So, the dynamic part of the TBM is related to belief revision (here belief transfer) following the awareness of new information.
The transfer of belief in the TBM satisfies the rule of conditioning of Dempster. Let B an event of 2^{Ω}, we consider for example that we have a new information which implies that all solutions of the problem are in B. The conditional mass $m[B](.)$ (the hooks represent conditioning) is given by not normalized rule of conditioning of Dempster according to equations (5):

$$\begin{aligned} &\text{if } A \subset B \text{ and } A \neq \phi: \\ &\quad m[B](A) = \frac{1}{1 - \sum_{C \subseteq \overline{B}} m(C)} \sum_{C \subseteq \overline{B}} m(A \cup C), \\ &\text{else } (A \not\subset B \text{ or } A = \phi): \\ &\quad m[B](A) = 0 \\ &\text{and } m[B](B) = 1 \end{aligned} \qquad (5)$$

3. PROPOSED APPROACH

The proposed classification system consists of three stages: First of all, the masses distribution calculation for each point of the image according to a classification based on spectral information, and then the beliefs transfer on the basis of contextual information (interior or external crown). Finally, a stage of fusion with the structural source of information provides a new distribution of combined masses.

3.1 Spectral classification based on belief functions theory

Classification by belief function theory requires, at the outset, an estimation of belief mass functions for the calculation of the resulting decision functions (credibility, plausibility or pignistic probability) on which the classification process decisions are based. We proposed in previous papers (Ben Dhiaf and all, 2007; Ben Dhiaf and all,2008-a), two methods of belief masses estimation based on grey levels histograms of learning zones. The first is a method that passes through a distribution of possibilities and the second directly reveals a belief mass estimation.

3.2 Integration of contextual information

In (Ben Dhiaf and all, 2008-b), we proposed a mean of conflict management by determining the subset of sources to use for each context. This approach reduces complexity since we consider, for each context, only a subset of sources validated by contextual variables. In this paper, we propose another method taking advantage of the context and allowing reducing the possible classes set for a given context.

Each context is determined according to contextual variables z_j and is described by a vector that we note *Context.* Each element α_i of the vector *Context* is a boolean value. It expresses if the class C_i is possible for this context or not. α_i = 1, if C_i is possible in the considered context, else $\alpha_i = 0$ (C_i belongs to an impossible class along with the considered context : $C_i \in \overline{Context}$).

Let $Z = \{z_1, z_2 \cdots, z_p\}$, the space of all the possible contexts, composed of p contexts z_j with $j \in \{1, \ldots, p\}$. We will describe each context with a row vector $Context = [\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_k]$, $\alpha_i \in \{0,1\}$. The size of this vector is equal to the cardinality of the frame of discernment Ω. Contexts can be written in the following form (6). Each row of the matrix corresponds to a context (determined according to contextual variables).

$$M\_contexts = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1k} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{p1} & \alpha_{p2} & \cdots & \alpha_{pk} \end{bmatrix} \quad (6)$$

Consider some examples of contextual variables: Altitude, Crown and Structure.

Altitude can take the values: high, average, low;

Crown can take the values: interior of a crown, exterior of a crown;

Structure can take the values: circular, rectangular, ellipsoidal.

In our application, we consider the contexts: Crown (interior of a crown) and $\overline{Crown}$ (exterior of a crown). Some classes (non forest classes) cannot belong to interior of a crown for example. The idea is then to transfer the mass associated with the impossible classes towards the possible classes.

We propose a transfer of beliefs according to the context crowns. We will call the masses obtained after beliefs transfer "contextual masses". We write the distribution of contextual masses in the following form (7) :

$$m(C_i) = x_i \quad (7)$$
$$0 < x_i \leq 1 \text{ if } C_i \in \text{crown } (context)$$
$$\text{and} \qquad x_i = 0 \text{ if } C_i \in \overline{\text{crown}} \ (\overline{context})$$

3.3 Integration of structural information

In this section we are interested in the integration of structural information in the fusion process. The indices of forms being able to be used are varied: area, perimeter, circularity, rectangularity, ellipticity. We retain for this application the area (surface) of the crowns as structural measure.

The distribution of mass that we propose for the structural source is inspired from distances calculations. Thus we write the structural mass of a crown as illustrated by equation (8):

$$m_{crown}(C_i) = \frac{1 - d_i}{D} \quad (8)$$

With d_i is the normalised distance (between areas) between the considered crown and the average area of the class C_i :

$$d_i = \frac{|a - a_i|}{norm} \text{ with } norm = \max_{i \in \{1,\ldots,k\}}(a_i) - \min_{i \in \{1,\ldots,k\}}(a_i)$$
$$D = \sum_{i \in \{1,\ldots,k\}}(d_i)$$

a : Area of the considered crown,

a_i : Average area of the crowns of C_i,

$\max_{i \in \{1,\ldots,k\}}(a_i)$: Maximum of the average areas of all classes,

and $\min_{i \in \{1,\ldots,k\}}(a_i)$: Minimum of the average areas of all the classes.

In the end of the proposed classification process, we propose to combine structural information relevant to a crown (described by the structural masses distribution) with the spectral information of pixels of the same crown (described by the mass distribution after transfer of beliefs based on contextual information: contextual masses distribution).

4. APPLICATION

This section illustrates the application of our approach on a window of the PIR band of high resolution Quickbird satellite image of an area at the north of Tunisia (Cf. figure 1). The forest inventory corresponding to the same scene of the image emphasizes four classes: Algerian oak, cork oak, naked soil, soil with little coverage.

We propose first to explain the general principle of the algorithm of trees crowns delimitation (extraction) by Brownian motion (paragraph. 4.1). After, we present results of spectral classification based on pignistic probabilities functions maximisation. The belief masses distribution is deduced from histograms of learning areas corresponding to classes of the image (paragraph. 4.2). Revision of this masses distribution on the basis of contextual information (cf. paragraph 3.2) provides a new one (paragraph 4.3). The last step of our approach allows combining with structural information (paragraph 4.4).

Figure 1. A window of the PIR band of the QuickBird image

4.1 Extraction of trees crowns by Brownian motion

The trees crowns extraction algorithm by Brownian motion (ECBM) can be divided into four steps: Pretreatments, extraction of local maxima, delimitation of trees crowns and definition of borders (Erickson, 2004).
As first pre-treatment, ECBM algorithm eliminates everything that is different from tree (naked soil, rocks... etc), then it calculates the distance to the background and performs a Gaussian smoothing.
The local maxima represent the tops of the trees. ECBM algorithm determines local maxima by application of a mask on the smoothed image.

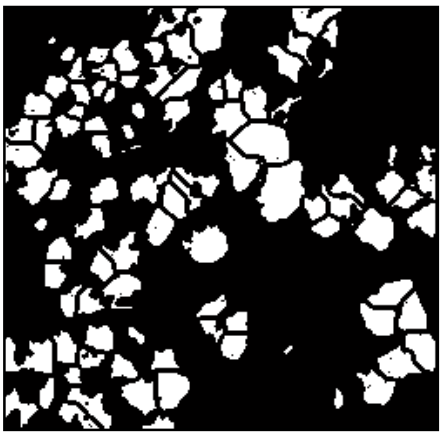


Figure 2. Extraction of trees crowns by Brownian motion algorithm

The phenomenon of the Brownian motion represents the random movement of a suspended particle in a fluid. Then, for each local maximum detected, the ECBM algorithm applies a Brownian movement to a particle initialized to the top to reach the crown of the tree. Position of the particle after N stages is equal to the sum of N random vectors of displacement of the particle. The limitation of the borders corrects the effects due to crowns overlapping. This limitation permits to obtain independent crowns ready to be classified.
The image of figure 3 is the binary image resulting of the application of the ECBM algorithm on the image of figure 1. Figure 3 illustrates the efficiency of this algorithm by superposing the limits of the detected crowns on the image of figure 1.

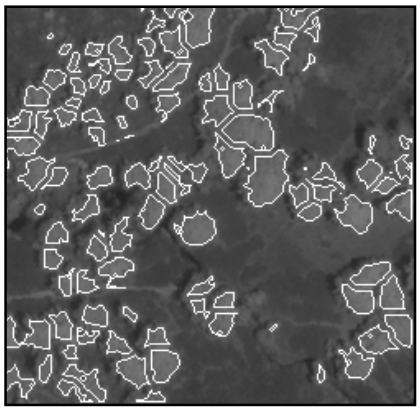
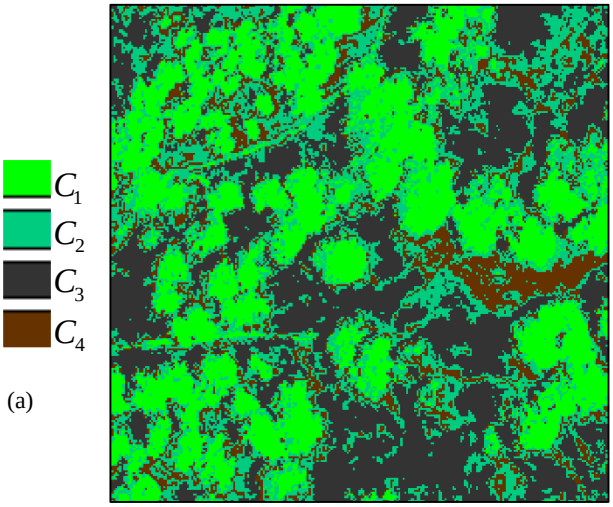


Figure 3. Result of matching of the Quickbird image window and the corresponding image of crowns

4.2 Spectral classification

Figure 4 illustrates the result of the image classification while being based on the maximum of the masses estimated on the basis of supervised training (Ben Dhiaf and all, 2007; Ben Dhiaf and all,2008-a),.

This result reveals a great confusion between forest and non forest species. Although the value of Kappa coefficient (0.81) and the mean of the values of confusion matrix diagonal (85.92) shows that our spectral classification isn't bad, a confusion essentially between class 2 is and the other classes 1 and 3 (cf. Table 5) needs to be reduced.



(a)

(b)

Figure 4. (a) : Legend of classes, (b) : Spectral classification based on maximization of pignistic probabilities functions.

	C1	C2	C3	C4
C1	92.00	8.00	0.00	0.00
C2	10.91	60.00	21.82	7.27
C3	0.00	3.28	96.72	0.00
C4	0.00	5.05	0.00	94.95

Table 5. Confusion matrix of the spectral classification

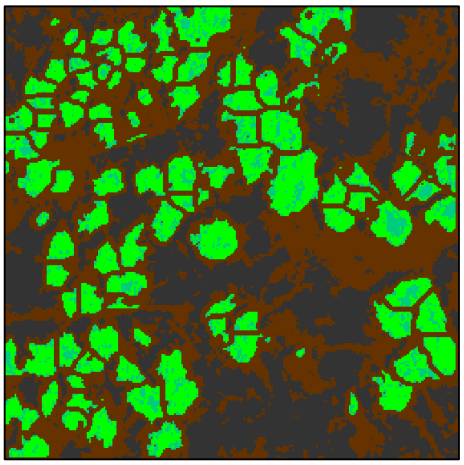


Figure 6: Result of classification after revision

4.3 Transfer of beliefs and new classification

Figure 7 illustrates the result of belief transfer applied on the basis of contextual information on the spectral mass distribution. We note that this transfer reduces considerably the presence of forest species pixels in classes C_3 and C_4 and concentrated the presence of the classes C_1 and C_2 in interior pixels of the trees crowns. At this step, the classification of figure 4 still leaves a great confusion between the two classes C_1 and C_2.

4.4 Fusion with structural information

Figure 8 shows the relevance of the integration of structural information (surface of the crowns). Indeed the figure shows a better distinction between the classes C_1 and C_2 in comparison with figure 3 and 4.

The following step consists on unification of classes assigned to pixels belonging to the same tree crown. The result shown in figure 8 converges to the field reality (comparison made relatively to the forest inventory and after discussion with the experts of the ministry for agriculture of Tunisia-direction of the forests). Table 9 presents an evaluation of tree crowns classification. The percentage of well classified tree crowns is satisfactory (94% for C1 and 88% for C2).

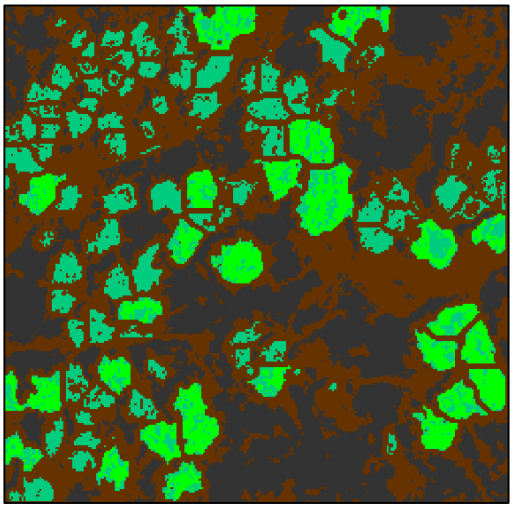


Figure 7: Classification after fusion with structural information

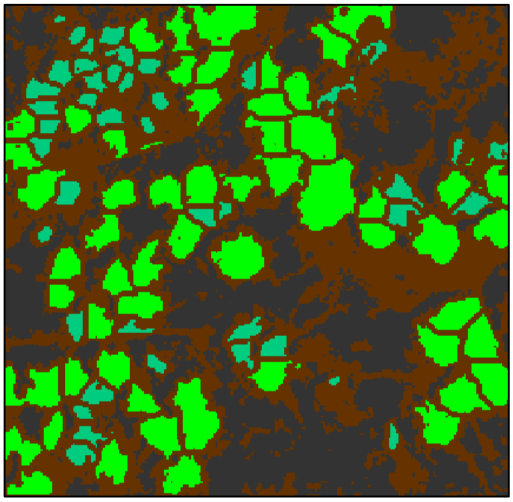


Figure 8: Classification after revision, combination and unifying the interior of crowns

	C1	C2
Tree crowns number	57	43
Misclassified crowns number	3	5
Well classified crowns number	54	38
Percentage of well classified crowns	94%	88%

Table 9: Evaluation of tree crowns classification

5. CONCLUSION

We presented in this paper an approach which makes it possible to integrate contextual information for the transfer of belief dedicated to the classification of the forest images. The taking into account of the context and expert knowledge in the revision of the masses enabled us to manage the conflict between the forest species. The results confirm well the importance of this choice. The advantage of this approach brings a double advantage: Separability between the forest species and a reduction of calculations complexity of the belief masses since contextual information enables us to filter combinations of classes not validated by the contextual assumptions.

REFERENCES

Z. Ben Dhiaf, J. Desachy and A. Hamouda, *Information sources fusion approach in forest stand classification*, SPIE, conference on "Image and Signal Processing for Remote Sensing", proceeding of SPIE vol. 67480V1-67480V10, Italy 17-21 September 2007.

Z. Ben Dhiaf, J. Desachy and A. Hamouda, *Management of the conflict between information sources in a fusion process: Application in forest cartography*, ISIVC'2008, 4th International Symposium on Image/Video Communications over fixed and mobile networks, Bilbao - Spain, July 9-11th.

Z. Ben Dhiaf, A. Hamouda et J. Desachy, *Théorie des fonctions de croyance : Fondements, complexité et application à la classification d'images*, Conférence Internationale sur relations, ordres et Graphes (ROGICS'08) : Interaction avec l'Informatique 12-17 Mai 2008 - Mahdia, Tunisie.

I.Bloch, Fusion d'informations numériques : panorama méthodologique, JNRR'O5. 7 octobre 2005.

I.Bloch, *Information Combination Operators for Data Fusion : A Comparative Review with Classification*, IEEE Transactions on Systems, Man, and Cybernetics, 26(1) : pp 52-67, Janvier 1996.

N.Burrus. Evidence Theory (part 1): Theoretical aspects. LRDE Seminar , Mai 2003.

T. Denoeux, Théorie des fonctions de croyance et classification, Séminaire IRIT, Toulouse, 2 juin 2004.

J. Desachy, H. Suzuki et V.Bessettes. "Pattern Recognition and Cooperation/Fusion Techniques: Some Practical Issues", conférence invitée. Image and Signal Processing for Remote Sensing VI, Barcelona, Spain, SPIE. Septembre 2000.

M. Erickson. *Segmentation and Classification of Individual Tree Crowns in High Spatial Resolution Aerial Images*. PhD thesis, Swedish University of Agricultural Sciences Uppsala, Sweden, 2004.

G. Shafer, *Perspectives on the theory and practice of belief functions*, International Journal of Approximate Reasoning, Vol. 4, Issue: 5-6, pages 323 – 362, 1990.

P. Smets, *The transferable belief model and other interpretations of Dempster-Shafer's model*, Uncertainty in Artificial Intelligence, 1990.

P.Vannoorenberghe, "Un état de l'art sur les fonctions de croyance appliquées au traitement de l'information", Revue Information, Interaction, Intelligence. I3, Volume 3, N° 2 pp 9-45 , 2003.

COMPARISON OF ERROR PROPAGATION IN BLOCK ORIENTATION: AN ANALYTICAL APPROACH

J. Cothren [a, *], B. Schaffrin[b]

[a]Department of Geosciences, University of Arkansas, Fayetteville, AR – jcothren@cast.uark.edu
[b]School of Earth Sciences, Ohio State University, Columbus, OH – schaffrin.1@osu.edu

KEY WORDS: Model, Orientation, Integration, Triangulation, Adjustment

ABSTRACT:

Many researchers have reported comparisons between the error propagation properties of direct orientation and indirect orientation (aerial triangulation using ground control points). The results of these comparisons have shown that direct orientation has the potential for use in projects requiring all but the highest accuracy. However, all of these empirical comparisons are specific to the particular configuration of the image block and sensor systems and do not provide an explicit analytical comparison of the general case. In this paper, we present an analytical comparison of the ground point precision obtained from direct and indirect orientation methods within the framework of a block-bundle adjustment in a stochastically constrained Gauss-Markov Model.

1. INTRODUCTION

Accurate and reliable sensor orientation is a pre-requisite for virtually all digital photogrammetric products including georeferenced orthoimages, digital terrain models and (increasingly common) three-dimensional models of man-made surface features. Until very recently analytical aerial triangulation (AT) was the most common method of sensor orientation. Modern methods of AT rely upon the bundle block adjustment, in which possibly large numbers of tie points are – often automatically via correlation or least-squares matching – measured across two or more images and used to estimate the six orientation parameters of each image in the block. Ground control points are integrated into the adjustment as weighted observations and provide datum information to accurately geo-reference the block. As Heipke et al. (2002) and others have remarked, AT is particularly advantageous because the control information resides near the ground features of interest and is therefore largely an interpolation problem. Exterior orientation parameters can be considered nuisance parameters. In fact, AT is often considered a ground control densification process**.

Not long after the Global Positioning System became operational, differential positioning technologies enabled the direct observation of camera exposure centers. The flexible bundle adjustment allowed these additional observations to be seamlessly integrated as weighted observations into the AT solution. This reduced, but did not eliminate, the need for ground control points for datum definition*** and georeferencing. However, tie points were still required for the estimation of the rotation matrix from each camera frame to the model frame.

In the late 1990's, Inertial Measurement Systems (IMU's) made it possible to directly observe a sensor's orientation relative to the ground. Direct observation of the exterior orientation using integrated GPS and IMU systems now allow sensor orientation without manual or automated measurement of either tie points or ground control points – albeit with potential reliability issues because of the lack of redundancy. Many of the original researchers – among them Skaloud and Schwarz (1998), Toth (1998), Burman (1999), Colomina (1999), Grejner-Brzezinska (1999), and Cramer et al. (2000) – demonstrated ground accuracies at the decimeter level without ground control. However, unlike AT solutions in which ground control points provided georeferencing information, these solutions involved extrapolation of control at the sensor to the ground. Furthermore, because the orientation parameters were no longer nuisance parameters, their correlation with each other and fixed interior orientation parameters could not be used advantageously by the adjustment to compensate for poor or variable calibration.

The extrapolation and reliability problems lead to the inclusion of all observations, GPS/IMU, ground control points and tie points into the bundle adjustment in a process popularly known as integrated sensor orientation, or ISO. In particular, as reported by Heipke et al. (2002), ISO may be used to more accurately estimate IMU boresight and GPS phase center offsets in a calibration step, but is also useful in a traditional AT role. In fact, the OEEPE Integrated Sensor Orientation tests (Heipke et al., 2002b) demonstrated mean deviations from independent check points of ± (5-10) cm in planimetry and ± (10-15) cm in height at a 1:5000 image scale. This compared to same-block accuracies from the AT solution of ± (2.0-2.8) cm in planimetry and ±3.2 cm in height. This, and other more recent experiences, clearly demonstrate that direct orientation, while inferior to more labor intensive AT in this case, is sufficient for many lower accuracy projects. At scales of 1:10,000 results showed a similar planimetric accuracy but ±7 cm height accuracy compared to ± (3 – 13.4) cm planimetric accuracy in the direct orientation solution. The wide range of results is due in part to the different system calibration parameters and adjustment coordinate frames.

* Corresponding author

** This was particularly true before widespread use of the bundle adjustment when tie points were used to transfer control throughout the block. Now it is more common to use the exterior orientation estimates produced by the bundle adjustment.

*** Jacobsen and Wegmann (2001) address issues regarding GPS observations in Cartesian frames.

The 2002 OEEPE tests were flown with a dual-frequency GPS receiver and very high quality IMU (Heipke et al., 2002). The tests do not, and were not intended to, shed light on the capability of other integrated sensor systems. With the variety of GPS/IMU configurations available, from low-cost systems to exceedingly capable and expensive systems (cf. Mostafa and Hutton, 2005), the need exists for an analytic relationship among the three orientation solutions – direct orientation, indirect orientation with ground control points, and integrated sensor orientation. The analytic solution should provide insight into the precision required of a pure direct orientation solution (with a boresight-calibrated GPS/IMU system) to achieve ground coordinate precisions equal to a similarly configured indirect AT solution. In this paper we derive inequalities within the framework of a constrained block bundle adjustment. While we do not address ISO explicitly (as did Habib and Schenk, 2002, though using a different approach), the developed framework is suitable for this orientation method as well. Section 2 describes the partitioned bundle block model constrained by pseudo-observations on the parameters (also known as "stochastic constraints"). Section 3 describes how orientation constraints enter the normal equations and affect parameter estimates. These results are then demonstrated at different image scales in Section 4 using a simulated numerical example. We discuss possible extensions to the analytical relationships in Section 5.

2. THE PARTITIONED BUNDLE BLOCK MODEL

The bundle block adjustment is the standard for photogrammetric aerial triangulation because of its comprehensive solution and flexibility. Observations consist of image tie-point measurements, independent ground control point measurements (2 image measurements per point and up to 3 ground coordinate measurements) and the direct observation of exterior orientation parameters with GPS/IMU measurements. If, at first, we only consider observations in the form of image measurements of ground control and tie points, an appropriate stochastic model for the linearized bundle block adjustment (unless additional distortion terms are carried as parameters) is the Gauss-Markov Model (GMM),

$$\mathbf{y} = \mathbf{A}\xi + \mathbf{e}\,, \qquad \mathrm{e} \sim (\mathbf{0}, \sigma_o^2 \mathbf{P}^{-1})\,, \tag{1}$$

in which

- **y** is a *n x 1* random vector of incremental changes to image coordinate observations,
- **A** is an *n x m* non-random design matrix of rank *q = m - 7*, representing the Jacobian matrix of the observation equations with respect to the unknown parameters thereby defining a local differential relationship between parameters and observations,
- ξ is the *m x 1* non-random incremental parameter vector of the linearized observation equations and
- **e** is the *n x 1* random error vector with first and second moments given. The weight matrix **P** is usually treated as a known value (typically it is assumed to be the identity matrix unless image point measurements are of different precision), and the variance component (or reference variance) σ_o^2 is considered as the scale factor for the variance of the image point measurements. If σ_o^2 is chosen to be 1 a priori, then the full observation covariance information is contained in $\mathbf{P}^{-1}$.

Note that in the absence of geo-referencing information from ground control points (i.e. indirect orientation) or direct observations of exterior orientation (i.e. direct orientation), **A** is rank-deficient and the multiple solutions for ξ represent the multiple coordinate frames in which the block can be positioned.

To reduce the computational requirements for solving large photogrammetric blocks, it is common practice to partition design matrices to achieve a particular sparse block-diagonal configuration (cf. Kraus, 1993). We follow this practice because it will later facilitate comparison between direct and indirect orientation, and partition the Gauss-Markov Model as follows:

$$\mathbf{y} = \mathbf{A}_1\xi_1 + \mathbf{A}_2\xi_2 + \mathbf{A}_3\xi_3 + \mathbf{e} \quad , \quad \mathbf{e} \sim (\mathbf{0}, \sigma_o^2 \mathbf{P}^{-1})\,, \tag{2}$$

in which

- $\mathbf{A}_1$ is a *n x* (*6 * number of photos*) matrix containing partial derivatives with respect to the exterior orientation parameters, and ξ_1 contains the incremental changes to the initial approximations of exterior orientation parameters;
- $\mathbf{A}_2$ is a *n* x (*3 * number of gcp points*) matrix containing partial derivatives with respect to the three coordinates (X,Y,Z) of the potential ground control points, and ξ_2 contains the incremental changes to the initial approximations of ground coordinates;
- $\mathbf{A}_3$ is a *n* x (*3 * number of tie points*) matrix containing the partial derivatives of the image coordinate observations with respect to the three ground coordinates of the tie points, and ξ_3 contains the incremental changes to the initial approximations of tie-point ground coordinates. Check points could be incorporated into this partition as well.

Absolute orientation information may enter the model through stochastic constraints on ξ_1 (direct orientation), ξ_2 (indirect orientation), or both. The stochastic constraints required for direct orientation:

$$\mathbf{z}_1 = \xi_1 + \mathbf{e}_1 \qquad \mathbf{e}_1 \sim \left(\mathbf{0}, \sigma_o^2 \mathbf{P}_1^{-1}\right), \tag{3}$$

provide additional information about the exterior orientation elements. Note that we use the same variance component as in (2) which may, for $\sigma_o^2 = 1$, imply that the full GPS/IMU observation covariance is contained in $\mathbf{P}_1^{-1}$. Likewise the stochastic constraints required by indirect orientation (observation of ground control points):

$$\mathbf{z}_2 = \xi_2 + \mathbf{e}_2 \qquad \mathbf{e}_2 \sim \left(\mathbf{0}, \sigma_o^2 \mathbf{P}_2^{-1}\right), \tag{4}$$

provide additional information about ground control point coordinates in ξ_2.

Since the quality of the triangulation is evaluated by the precision (and accuracy) of points on the ground, one measure of the relative quality achieved by each orientation method may be evaluated by the post-adjustment covariance matrix, $\mathbf{Q}_3$, of the estimated tie-point coordinates in ξ_3. In the next section we explicitly express this covariance matrix in terms of both $\mathbf{P}_1$ and $\mathbf{P}_2$. We assume that tie points are measured automatically in the

direct method (but with no control points measured) and in the indirect method (with measured control points). $\mathbf{A}_2$ and $\mathbf{A}_3$ are therefore invariant with respect to the method of georeferencing (and thus to the added stochastic constraints). Furthermore, in the direct method, the ground coordinate parameters contained in ξ_2 are treated as unknowns (i.e. as tie points).

3. CONSTRAINED NORMAL EQUATIONS OF THE PARTITIONED MODEL

The partitioned rank-deficient normal equations for (2) are obtained as

$$\begin{bmatrix} \mathbf{N}_{11} & \mathbf{N}_{12} & \mathbf{N}_{13} \\ \mathbf{N}_{21} & \mathbf{N}_{22} & \mathbf{0} \\ \mathbf{N}_{31} & \mathbf{0} & \mathbf{N}_{33} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\xi}}_1 \\ \hat{\boldsymbol{\xi}}_2 \\ \hat{\boldsymbol{\xi}}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \mathbf{c}_3 \end{bmatrix} \tag{5}$$

with $\left[\mathbf{N}_{ij}, \mathbf{c}_i\right] = \mathbf{A}_i^T \mathbf{P}\left[\mathbf{A}_j^T, \mathbf{y}\right]$, and may be augmented with (3) to provide absolute orientation information through direct orientation

$$\begin{bmatrix} \mathbf{N}_{11}+\mathbf{P}_1 & \mathbf{N}_{12} & \mathbf{N}_{13} \\ \mathbf{N}_{21} & \mathbf{N}_{22} & \mathbf{0} \\ \mathbf{N}_{31} & 0 & \mathbf{N}_{33} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\xi}}_1 \\ \hat{\boldsymbol{\xi}}_2 \\ \hat{\boldsymbol{\xi}}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{c}_1+\mathbf{P}_1\mathbf{z}_1 \\ \mathbf{c}_2 \\ \mathbf{c}_3 \end{bmatrix}, \tag{6}$$

or augmented with (4) to provide absolute orientation information via indirect orientation

$$\begin{bmatrix} \mathbf{N}_{11} & \mathbf{N}_{12} & \mathbf{N}_{13} \\ \mathbf{N}_{21} & \mathbf{N}_{22}+\mathbf{P}_2 & \mathbf{0} \\ \mathbf{N}_{31} & \mathbf{0} & \mathbf{N}_{33} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\xi}}_1 \\ \hat{\boldsymbol{\xi}}_2 \\ \hat{\boldsymbol{\xi}}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2+\mathbf{P}_2\mathbf{z}_2 \\ \mathbf{c}_3 \end{bmatrix} \tag{7}$$

Note that the zero blocks in the normal equations (5)-(7) indicate that no observation equations contain both tie and control point coordinates. This is strictly true only in the absence of additional observations in the form of, for example, known distances between a tie and control point. In both cases the addition of sufficient stochastic constraints resolves the rank-deficiency of the normal equations in (5). However, the precision with which the coordinate estimates of the tie-points, $\hat{\boldsymbol{\xi}}_3$, are determined depends upon the structure of the normal equations. Again, since we are concerned with the precision of the ground coordinates in the triangulation, we will examine the effects of (6) versus (7) on the dispersion of $\hat{\boldsymbol{\xi}}_3$. The covariance matrix (or cofactor matrix, if we do not use the a-priori value for the variance component) of all adjusted parameters is contained in a generalized inverse of the normal equations matrix. It can be shown that a reflexive, symmetric generalized inverse of the normal matrix in (5) is given by

$$\begin{bmatrix} \mathbf{N}_{11} & \mathbf{N}_{12} & \mathbf{N}_{13} \\ \mathbf{N}_{21} & \mathbf{N}_{22} & \mathbf{0} \\ \mathbf{N}_{31} & \mathbf{0} & \mathbf{N}_{33} \end{bmatrix}^- = \begin{bmatrix} \mathbf{S}^- & -\mathbf{S}^-\mathbf{N}_{12}\mathbf{N}_{22}^{-1} & -\mathbf{S}^-\mathbf{N}_{13}\mathbf{N}_{33}^{-1} \\ -\mathbf{N}_{22}^{-1}\mathbf{N}_{21}\mathbf{S}^- & \mathbf{N}_{22}^{-1}+\mathbf{N}_{22}^{-1}\mathbf{N}_{21}\mathbf{S}^-\mathbf{N}_{12}\mathbf{N}_{22}^{-1} & \mathbf{N}_{22}^{-1}\mathbf{N}_{21}\mathbf{S}^-\mathbf{N}_{13}\mathbf{N}_{33}^{-1} \\ -\mathbf{N}_{33}^{-1}\mathbf{N}_{31}\mathbf{S}^- & \mathbf{N}_{33}^{-1}\mathbf{N}_{31}\mathbf{S}^-\mathbf{N}_{12}\mathbf{N}_{22}^{-1} & \mathbf{N}_{33}^{-1}+\mathbf{N}_{33}^{-1}\mathbf{N}_{31}\mathbf{S}^-\mathbf{N}_{13}\mathbf{N}_{33}^{-1} \end{bmatrix} \tag{8}$$

with $\mathbf{S} = \mathbf{N}_{11} - \mathbf{N}_{12}\mathbf{N}_{22}^{-1}\mathbf{N}_{21} - \mathbf{N}_{13}\mathbf{N}_{33}^{-1}\mathbf{N}_{31}$. Here $\mathbf{S}^-$ denotes any reflexive, symmetric g-inverse of $\mathbf{S}$ (Magnus and Neudecker, 1999, pg. 12). The lower right block is the covariance matrix of the tie-point coordinate estimates and, thus, the matrix that the users would like to see minimized. The pseudo-observations, whether from direct or indirect information, affect $\mathbf{Q}_3$ through $\mathbf{S}$. If geo-referencing information is provided directly through calibrated GPS/IMU observations, the pseudo-observations in (6) create the, now full-rank, matrix

$$\mathbf{S}_1 = \mathbf{N}_{11} + \mathbf{P}_1^{-1} - \mathbf{N}_{12}\mathbf{N}_{22}^{-1}\mathbf{N}_{21} - \mathbf{N}_{13}\mathbf{N}_{33}^{-1}\mathbf{N}_{31} \tag{9}$$

instead of $\mathbf{S}$, and we may write the tie-point cofactor matrix in terms of the inverse of $\mathbf{S}_1$ as

$$\mathbf{Q}_3^{(1)} = \mathbf{N}_{33}^{-1} + \mathbf{N}_{33}^{-1}\mathbf{N}_{31}\mathbf{S}_1^{-1}\mathbf{N}_{13}\mathbf{N}_{33}^{-1}. \tag{10}$$

The superscript $^{(1)}$ is used to indicate that the tie point coordinate estimates are obtained via the direct method. Alternatively, if the georeferencing information is provided indirectly, through ground control points only, the pseudo-observations in (7) create the full-rank matrix

$$\mathbf{S}_2 = \mathbf{N}_{11} - \mathbf{N}_{12}\left(\mathbf{N}_{22}+\mathbf{P}_2\right)^{-1}\mathbf{N}_{21} - \mathbf{N}_{13}\mathbf{N}_{33}^{-1}\mathbf{N}_{31}. \tag{11}$$

The covariance matrix of the tie-point coordinates achieved through indirect orientation may thus be written as

$$\mathbf{Q}_3^{(2)} = \mathbf{N}_{33}^{-1} + \mathbf{N}_{33}^{-1}\mathbf{N}_{31}\mathbf{S}_2^{-1}\mathbf{N}_{13}\mathbf{N}_{33}^{-1} \tag{12}$$

In this case the superscript $^{(2)}$ indicates that the estimates were obtained via the indirect method. Now we are ready to analytically relate the tie-point precision to the orientation information contained in $\mathbf{P_1}$ or $\mathbf{P_2}$, respectively. If $\mathbf{Q}_3^{(1)} \le \mathbf{Q}_3^{(2)}$, then direct orientation leads to a better tie-point precision than indirect orientation; for details of this comparison see Marshall and Olkin (1979). From (10) and (12) we see that this inequality holds *if and only if* $\mathbf{S_1} \ge \mathbf{S_2}$. This is to be expected since from (8) we know that the inverses of $\mathbf{S_1}$ and $\mathbf{S_2}$ are the respective covariance matrices of the exterior orientation elements if constrained by direct and indirect observations, respectively. By substituting (9) and (11) into this inequality it can be shown (see Appendix A for details) that

$$\mathbf{Q}_3^{(1)} \le \mathbf{Q}_3^{(2)} \Leftrightarrow \mathbf{P}_1 \ge \mathbf{N}_{12}\mathbf{N}_{22}^{-1}\left(\mathbf{N}_{22}^{-1}+\mathbf{P}_2^{-1}\right)^{-1}\mathbf{N}_{22}^{-1}\mathbf{N}_{21}. \tag{13}$$

This expression provides us with a measure to determine the weight matrix (or the covariance matrix) of the direct stochastic constraints that is required to achieve, at least, the same precision as an indirect adjustment with a given ground control

configuration (which is contained in $\mathbf{N}_{22}$ and $\mathbf{N}_{12}$) and precision (contained in $\mathbf{P}_2$). Continued manipulation of (13) can likewise isolate $\mathbf{P}_2$, leading to the equivalence

$$\mathbf{Q}_3^{(1)} \le \mathbf{Q}_3^{(2)} \Leftrightarrow$$

$$\mathbf{P}_2 \le \mathbf{N}_{22}\left(\begin{array}{c}\mathbf{N}_{21}\mathbf{N}_{11}^{-1}\mathbf{N}_{12}\left(\mathbf{N}_{21}\mathbf{N}_{11}^{-1}\mathbf{P}_1\mathbf{N}_{11}^{-1}\mathbf{N}_{21}\right)^{-1}\cdot \\ \cdot\ \mathbf{N}_{21}\mathbf{N}_{11}^{-1}\mathbf{N}_{21}-\mathbf{N}_{22}\end{array}\right)^{-1}\mathbf{N}_{22}. \qquad (14)$$

The inequalities (13) and (14) are satisfied when direct orientation provides uniformly higher precision estimates of the tie point coordinates. It is important to note that the tie point configuration, which is assumed to be invariant between the two geo-referencing methods, does not affect the comparison. Therefore, these inequalities also hold for any number of tie points included in the adjustment. Furthermore, ground point coordinates computed by intersection apart from the bundle adjustment (as in DEM generation and feature extraction) benefit from the chosen orientation method quite similarly to tie points used in the adjustment. This is true because $\mathbf{S}_1$ and $\mathbf{S}_2$ (the weight matrices of the estimated exterior orientation parameters) are used directly in the estimation of the precision of subsequent ground points. Also, we assume that the image coordinates of the ground control points are included in the adjustment, regardless of the orientation method, and simply serve as additional tie points in the direct method if they appear in multiple images.

More details about the matrix inequalities used throughout the preceding analysis can be found in Appendix B. In the next section we use the key results (13) and (14) in an example involving simulated fifty-six-image aerial blocks at two different scales.

4. COMPARISON IN A SIMULATED BLOCK: AN EXAMPLE

To demonstrate how the analysis above might be practically used, we consider two blocks, each composed of four strips of seven images each. One block is to be acquired at a scale of 1:4,800, the other at a sale of 1:24,000. Details of each block are given in Table 1 and the network configuration is shown in Figure 2. The ground control points for both blocks are assumed to be collected with GPS techniques yielding a horizontal coordinate variance of 0.0025 m^2 and a vertical coordinate variance of 0.0225 m^2 with (assumed) zero correlation among the coordinates. Also, the exposure center coordinates and orientation parameters are assumed to be observed with GPS/IMU devices yielding coordinate variances of 0.01 m^2 in the horizontal and 0.09 m^2 in the vertical, again without correlation. Orientation angles can be observed with a variance of 0.0001 $degrees^2$ in omega and phi, 0.01 $degrees^2$ in kappa with zero correlation. These variances reflect results achieved from calibrated systems (Heipke et al., 2002).

Block	Over –lap (%)	Side-lap (%)	B/H Ratio	Focal Len. (mm)	Format (mm)
1:4,800	60	30 %	1:1.6	150	235 x 235
1:24,000	60	30 %	1:1.6	150	235 x 235

Table 1. Acquisition parameters of the simulated block.

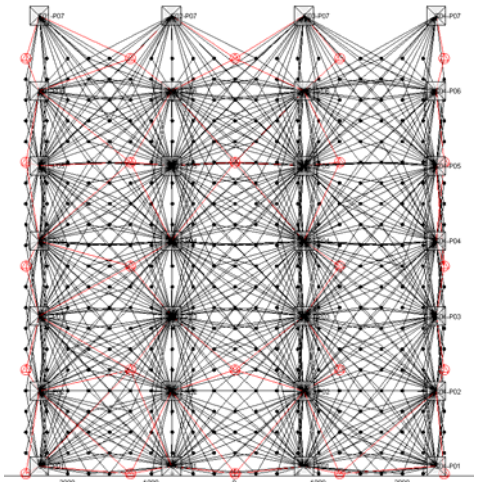


Figure 2. Configuration of the simulated block. Ground points used as control are shown in red.

A simulated trial consists of an aerial image block generated from the "true" values of both exterior orientation and ground points (both tie and control) as follows:

Direct orientation, simulated dataset:

1. Image point coordinates (x,y) are generated by perturbing the "true" values (generated by the collinearity equations using the "true" values of exterior orientation and ground points) with normally distributed random errors with a standard deviation of ±0.005 mm.
2. Initial estimates of the exterior orientation parameters are generated by perturbing the "true" values with normally distributed random errors with a standard deviation consistent with the assumptions in the preceding paragraph.
3. Initial estimates of the ground point coordinates (including ground control points) are intersected using the image coordinates developed in the first step and the exterior orientation values developed in the second step.
4. Pseudo-observations are added to the exterior orientation parameters per equation(3).

Indirect orientation, simulated dataset:

1. Image point coordinates (x,y) are generated as in the direct case.
2. Ground control point coordinates are perturbed with normally distributed random errors with a standard deviation consistent with the assumptions in the preceding paragraph.
3. Initial estimates for the horizontal coordinates of the exposure station are generated by solving a two-dimensional similarity transform between image coordinates and ground control points (2 in each image of the block). Exposure station Z-coordinate is assigned the flying height. The orientation angles phi and omega are assumed to be zero, kappa is assumed zero degrees in strips 1 and 3 and 180 degrees in strips 2 and 4.
4. Initial estimates of the tie points are intersected using the image coordinates developed in step one and the exterior orientation parameters developed in step 3.
5. Pseudo-observations are added to the ground control points per equation (4).

These results show that networks of a larger scale tend to yield comparable accuracies between the two orientation procedures, although horizontal accuracy appears to improve substantially with indirect orientation. The results also confirm our expectations that, as the photo scale decreases, both accuracy and precision are more drastically improved by indirect orientation procedures as compared to direct orientation procedures. Note that in both blocks the difference between the horizontal accuracies is somewhat larger than the difference between the vertical accuracies.

	Direct (±m)	Indirect (±m)	Direct (±m)	Indirect (±m)
Horz RMS	0.18	0.09	0.48	0.16
Total RMS	0.27	0.26	0.69	0.31
$\|\Delta X\|$	0.13	0.06	0.28	0.10
$\|\Delta Y\|$	0.10	0.05	0.33	0.11
$\|\Delta Z\|$	0.23	0.17	0.43	0.24
$3\sigma_X$	0.33	0.21	1.03	0.49
$3\sigma_Y$	0.33	0.21	1.12	0.47
$3\sigma_Z$	0.68	0.61	1.70	1.16

Table 3. Mean tie point coordinate results of 100 trials. RMS values compare estimated tie point coordinates with their "true" values. Sigma values are derived from variances propagated through the orientation adjustment.

From the point of view of the network designer, equation (13) is most useful in that it tells what modifications to the direct observations of the exterior orientation parameters are required to achieve the precision of the indirect method. We, therefore, consider the normal equations formed from the "design" values of the tie points, ground control points, and exterior orientation used to generate the simulations. The matrix norms of the partitioned normal equations and pseudo-observation weight matrices are shown in Table 3. Note that from these measurements we may determine the decrease of variance of each parameter estimate (the entries in the diagonal matrix $\mathbf{P}_1$) required to achieve precisions equal to orientation with ground control points with variances reflected in $\mathbf{P}_2$. According to row 3 in Table 3, the exterior orientation parameters would have to be observed with, on average, 1.9 times higher precision in the 1:4,800 scale block, and with 5.4 times higher precision in the 1:24,000 scale block.

Although each trial is developed from the same true values, each is unique due to the added random perturbations. Both the direct and indirect trials are consistent with standard operational procedures for providing initial approximations. The mean results of 100 trials are detailed in Table 3. All RMS values are considered accuracy measures in that they show mean deviations from the "true" values. The sigma values are propagated errors form the orientation procedures.

Partitioned matrices (eq.14)	Frobenius Norm 1:4,800	Frobenius Norm 1:24,000
$\mathbf{P_1}$	92.8×10^6	92.8×10^6
Eq. 13	338.0×10^6	2688.0×10^6
$\mathbf{P_2}$	283	283
Eq. 14	4690	493

Table 4. Matrix norms of the partitioned matrices.

	1:4,800	1:24,000
Frobenius Norm P_1	9.2×10^7	9.2×10^7
Frobenius Norm RHS (14)	3.37×10^8	2.687×10^9
P_1 scaling required	3.640	28.945
Direct Observation Variance Decrease Factor	0.275	0.0345

Table 5. Scaling of the weight matrix for direct orientation required to achieve a precision equal to the indirect method

5. CONCLUSIONS AND OUTLOOK

The framework developed in this paper has been used to derive an analytic relationship between two differently constrained bundle adjustments – direct and indirect – and is a step towards a more general method to compare various weighting methods for bundle adjustments. The framework may also be extended to include 1) the ISO method and 2) and integrated LiDAR observations, thereby following similar lines as Burman (2000) and Csanyi and Toth (2007).

REFERENCES

Burman, H., 1999. Using GPS and INS for orientation of aerial photography, *Proceedings of the ISPRS Workshop "Direct versus indirect methods of sensor orientation"*, Barcelona, pp. 148-157.

Burman, H., 2000. Adjustment of laser scanner data for correction of orientation errors. *Intl. Arch. of Photogrammetry and Remote Sensing*, v. 33, Part B3, pp. 125.132.

Caspary, W. F., 1987. *Concepts of Network and Deformation Analysis.* Monograph 11, School of Geomatics Engineering (formerly Surveying), The University of New South Wales, Kensington, NSW, Australia.

Colomina, I., 1999. GPS/INS and Aerial Triangulation: What is the best way for the operational determiniation of photogrammetric image orientation? *Intl. Arch. of Photogrammetry and Remote Sensing*, v. 32, Part 3-2W5, pp. 121.-130.

Cramer, M., Stallmann, D. and Haala, N., 2000. Direct georeferencing using GPS/inertial exterior orientations for photogrammetric applications; *Intl. Arch. of Photogramm. and Remote Sensing*, v. 33, part B3, pp. 198-205.

Csanyi, N., and Toth, C., 2007. Improvement of LiDAR data accuracy using LiDAR specific ground targets. *Photogramm. Engrg. & Remote Sensing*, 73(4), pp. 385-396.

Grejner-Brzezinska, D.A., 1999. Direct Exterior Orientation of Airborne Imagery with GPS/INS System: Performance Analysis. *Navigation*, Vol. 46, No. 4, pp. 261-270.

Grejner-Brzezinska, D.A., 2001. Direct Sensor Orientation in Airborne and Land-based Mapping Applications. *Geodetic & GeoInformation Science Report No. 461*. Department of Civil and Environmental Engineering and Geodetic Science. The Ohio State University, Columbus, OH, USA.

Habib, A., and Schenk, T., 2002. Accuracy analysis of reconstructed points in object space from direct and indirect

exterior orientation methods, In: *OEEPE Official Publ.* No 43, pp. 47-52.

Heipke C., Jacobsen, K., and Wegmann, H, 2002. Analysis of the results of the OEEPE test of Integrated Sensor Orientation. In: *OEEPE Official Publ*. No 43, pp. 31-45.

Jacobsen, K., and Wegmann, H., 2001. Dependencies and problems of direct sensor orientation. In: *OEEPE Official Publ.* No 43, pp. 73-84.

Kraus, K., 1993. *Photogrammetry, Volume 1, Fundamentals and Standard Processes.* Duemmler: Bonn, pp. 279-289.

Magnus, J.R., and Neudecker, H., 1999. *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Wiley: Chichester, pp. 11-12.

Marshall, A.W., and Olkin, I., 1979. Albert W. Marshall, Ingram Olkin, *Inequalities: Theory of Majorization and its Applications*. *Mathematics in Science and Engineering*, Vol. 143, Academic Press, New York, 1979, 569 pages.

Mostafa, M.M.R., and Hutton, J., 2005. A Fully Integrated Solution for Aerial Surveys: Design, Development, and Performance Analysis. *Photogramm. Engrg. & Remote Sensing*, 71(4), pp. 391-398.

Skaloud, J., and Schwartz, K.P., 1998. Accurate orientation for airborne mapping systems. *Intl. Arch. of Photogramm. and Remote Sensing*, v. 32, Part 2, pp. 282-290.

Toth, C., 1998. Direct platform orientation of multi-sensor data acquisition systems. *Intl. Arch. of Photogramm. and Remote Sensing*, v. 32, Part 4, pp. 629-634.

APPENDIX A. DERIVATION OF COVARIANCE RELATIONS

For the direct method to achieve equivalent or better ground point accuracies, $\mathbf{S}_1 \geq \mathbf{S}_2$. Using expressions (13) and (14) then

$$\mathbf{N}_{11} + \mathbf{P}_1 - \mathbf{N}_{12}\mathbf{N}_{22}^{-1}\mathbf{N}_{21} - \mathbf{N}_{13}\mathbf{N}_{33}^{-1}\mathbf{N}_{31} \geq$$
$$\mathbf{N}_{11} - \mathbf{N}_{12}\left(\mathbf{N}_{22} + \mathbf{P}_2\right)^{-1}\mathbf{N}_{21} - \mathbf{N}_{13}\mathbf{N}_{33}^{-1}\mathbf{N}_{31}$$

must hold with respect to the precision of the competing orientation methods. We may obviously eliminate the term containing tie-point information because it is invariant with respect to the orientation method and obtain

$$\mathbf{P}_1 - \mathbf{N}_{12}\mathbf{N}_{22}^{-1}\mathbf{N}_{21} \geq -\mathbf{N}_{12}\left(\mathbf{N}_{22} + \mathbf{P}_2\right)^{-1}\mathbf{N}_{21}$$
$$\mathbf{P}_1 \geq \mathbf{N}_{12}\mathbf{N}_{22}^{-1}\mathbf{N}_{21} - \mathbf{N}_{12}\left(\mathbf{N}_{22} + \mathbf{P}_2\right)^{-1}\mathbf{N}_{21}$$
$$\mathbf{P}_1 \geq \mathbf{N}_{12}\mathbf{N}_{22}^{-1}\mathbf{N}_{21} - \mathbf{N}_{12}\mathbf{N}_{22}^{-1}\mathbf{N}_{21} + \mathbf{N}_{12}\mathbf{N}_{22}^{-1}\left(\mathbf{N}_{22}^{-1} + \mathbf{P}_2^{-1}\right)^{-1}\mathbf{N}_{22}^{-1}\mathbf{N}_{21}$$

$$\mathbf{P}_1 \geq \mathbf{N}_{12}\mathbf{N}_{22}^{-1}\left(\mathbf{N}_{22}^{-1} + \mathbf{P}_2^{-1}\right)^{-1}\mathbf{N}_{22}^{-1}\mathbf{N}_{21}$$
$$\mathbf{P}_1 \geq \mathbf{N}_{12}\mathbf{N}_{22}^{-1}\left(\mathbf{I} + \mathbf{N}_{22}\mathbf{P}_2^{-1}\right)\mathbf{N}_{21}$$
$$\mathbf{P}_1 \geq \mathbf{N}_{12}\left(\mathbf{N}_{22} + \mathbf{N}_{22}\mathbf{P}_2^{-1}\mathbf{N}_{22}\right)^{-1}\mathbf{N}_{21}.$$

If $\mathbf{N}_{21}$ has full-row rank, then we may further rearrange the expression to obtain the following equivalent statements

$$\mathbf{N}_{21}\mathbf{N}_{11}^{-1}\mathbf{P}_1\mathbf{N}_{11}^{-1}\mathbf{N}_{12} \geq \mathbf{N}_{21}\mathbf{N}_{11}^{-1}\mathbf{N}_{12}\left(\mathbf{N}_{22} + \mathbf{N}_{22}\mathbf{P}_2^{-1}\mathbf{N}_{22}\right)^{-1}\mathbf{N}_{21}\mathbf{N}_{11}^{-1}\mathbf{N}_{12}$$
$$\left(\mathbf{N}_{21}\mathbf{N}_{11}^{-1}\mathbf{N}_{12}\right)^{-1}\mathbf{N}_{21}\mathbf{N}_{11}^{-1}\mathbf{P}_1\mathbf{N}_{11}^{-1}\mathbf{N}_{12}\left(\mathbf{N}_{21}\mathbf{N}_{11}^{-1}\mathbf{N}_{12}\right)^{-1} \geq \left(\mathbf{N}_{22} + \mathbf{N}_{22}\mathbf{P}_2^{-1}\mathbf{N}_{22}\right)^{-1}$$
$$\mathbf{N}_{22} + \mathbf{N}_{22}\mathbf{P}_2^{-1}\mathbf{N}_{22} \geq \left(\mathbf{N}_{21}\mathbf{N}_{11}^{-1}\mathbf{N}_{12}\right)\left(\mathbf{N}_{21}\mathbf{N}_{11}^{-1}\mathbf{P}_1\mathbf{N}_{11}^{-1}\mathbf{N}_{12}\right)^{-1}\left(\mathbf{N}_{21}\mathbf{N}_{11}^{-1}\mathbf{N}_{12}\right)$$
$$\mathbf{P}_2^{-1} \geq \left(\mathbf{N}_{22}^{-1}\mathbf{N}_{21}\mathbf{N}_{11}^{-1}\mathbf{N}_{12}\right)\left(\mathbf{N}_{21}\mathbf{N}_{11}^{-1}\mathbf{P}_1\mathbf{N}_{11}^{-1}\mathbf{N}_{12}\right)^{-1}\left(\mathbf{N}_{21}\mathbf{N}_{11}^{-1}\mathbf{N}_{21}\mathbf{N}_{22}^{-1}\right) - \mathbf{N}_{22}^{-1}$$

leading finally to

$$\mathbf{P}_2 \leq \mathbf{N}_{22}\left(\begin{array}{c}\mathbf{N}_{21}\mathbf{N}_{11}^{-1}\mathbf{N}_{12}\left(\mathbf{N}_{21}\mathbf{N}_{11}^{-1}\mathbf{P}_1\mathbf{N}_{11}^{-1}\mathbf{N}_{12}\right)^{-1}\cdot \\ \cdot\ \mathbf{N}_{21}\mathbf{N}_{11}^{-1}\mathbf{N}_{12} - \mathbf{N}_{22}\end{array}\right)^{-1}\mathbf{N}_{22}$$

APPENDIX B. MATRIX NORMS AND LÖWNER'S PARTIAL ORDERING OF MATRICES

Given two matrices of equal size, **G** and **H**, a partial ordering according to Löwner (cf. Marshall and Olkin,1979) can be defined through $\mathbf{G} > \mathbf{H}$ if and only if $\mathbf{G} - \mathbf{H}$ is positive-definite, or slightly more generally, $\mathbf{G} \geq \mathbf{H}$ if and only if $\mathbf{G} - \mathbf{H}$ is positive-semi-definite. In the case that **G** itself is positive-definite and **H** is, at least, positive-semi-definite, we may conclude $\mathbf{G} \geq \mathbf{H}$ if and only if the maximum eigenvalue of the positive-semi-definite matrix $\mathbf{G}^{-1}\mathbf{H}$ is less than one.

A justification for this is related to the relative size of a quadratic function of the two matrices. For details see Caspary (1987). Instead of comparing the matrices themselves, we may compare certain scalar-valued functions of them; for instance, the trace of **G** with the trace of **H**. Another possibility is the (weighted-) Frobenius norm, defined as $\|\mathbf{G}\|_w^2 = tr\left(\mathbf{G}^T\mathbf{W}\mathbf{G}\right)$ for some positive-definite matrix, **W**. For $\mathbf{W} \equiv \mathbf{I}$, this norm is equal to the sum of squared eigenvalues of **G** and can be thought of as a measure of the "hyper-volume" of the positive semi-definite matrix **G**. The weighted Frobenius norm provides a simple composite measure that can be used to scale the norm of **G** to that of **H**. For example, if $\frac{\|\mathbf{G}\|_w}{\|\mathbf{H}\|_w} = r \leq 1$ then by this measure at least, **G** is "smaller" than **H** by a factor of *r*. *r* may be applied as a scale factor to the norm of **G** then, $r\sqrt{tr\left(\mathbf{G}^T\mathbf{G}\right)} = \sqrt{tr\left(r\mathbf{G}^T r\mathbf{G}\right)} = \sqrt{tr\left(\bar{\mathbf{G}}^T\bar{\mathbf{G}}\right)}$ so that the norm of matrix $\bar{\mathbf{G}}$ is now equal to that of **H**.

A REVIEW ON IMAGE SEGMENTATION TECHNIQUES WITH REMOTE SENSING PERSPECTIVE

V. Dey [a, *], Y. Zhang [a], M. Zhong [b]

[a] Department of Geodesy and Geomatics Engineering, University of New Brunswick (UNB), Fredericton, E3B 5A3, NB, Canada – (d1991, yunzhang)@unb.ca
[b] Department of Civil Engineering, UNB, Fredericton, E3B 5A3, NB, Canada – ming@unb.ca

KEY WORDS: Image, Segmentation, Model, Measurement, Optical

ABSTRACT:

With the growing research on image segmentation, it has become important to categorise the research outcomes and provide readers with an overview of the existing segmentation techniques in each category. In this paper, different image segmentation techniques applied on optical remote sensing images are reviewed. The selection of papers include sources from image processing journals, conferences, books, dissertations and thesis out of more than 3000 journals, books and online research databases available at UNB. The conceptual details of the techniques are explained and mathematical details are avoided for simplicity. Both broad and detailed categorisations of reviewed segmentation techniques are provided. The state of art research on each category is provided with emphasis on developed technologies and image properties used by them. The categories defined are not always mutually independent. Hence, their interrelationships are also stated. Finally, conclusions are drawn summarizing commonly used techniques and their complexities in application

1. INTRODUCTION

Image segmentation in general is defined as a process of partitioning an image into homogenous groups such that each region is homogenous but the union of no two adjacent regions is homogenous (Pal and Pal, 1993). Efficient image segmentation is one of the most critical tasks in automatic image processing (Pavlidis, 1988; Haralick and Shapiro, 1985; Pal and Pal, 1993; Zhang, 1997; Cheng et al., 2001). Image segmentation has been interpreted differently for different applications. For example, in machine vision applications, it is viewed as a bridge between low level and high level vision subsystems (SpirKovska, 1993), in medical imaging as a tool to delineate anatomical structure and other regions of interest whose a priori knowledge is generally available (Pham et al., 2000) and in statistical analysis, it is posed as a stochastic estimation problem, with assumed prior distributions on image structure, which is widely used in remote sensing (Kerfoot et al., 1999). In remote sensing, it is often viewed as an aid to landscape change detection and land use/cover classification. Aforementioned examples state that image segmentation is present in every kind of image analysis. This constitutes a plethora of literature on the image segmentation. This necessitates the organized categorisation of them. In order to present an organized review on image segmentation techniques, this review paper limits its analysis to optical remote sensing image analysis. This is essential because radar image segmentation is another horizon in remote sensing image analysis. From now onwards, remote sensing image would refer only to optical satellite remote sensing images.

Optical remote sensing imagery has been to a paradigm shift in the decade after year 1999. Landsat 7 launched in 1999 (with Multispectral (MS), 30m spatial resolution; Panchromatic (Pan), 15m spatial resolution), IKONOS launched in 1999 (MS, 4.0m; Pan, 1.0m), Quickbird launched in 2001 (MS, 2.44m; Pan, 0.61m), WorldView-1 launched in 2007 (Pan, 0.5m), GeoEye-1 launched in 2008 (MS, 1.65m; Pan, 0.42m), and WorldView-2 launched in 2009 (MS, 1.8m; Pan, 0.46m) are evidence of this shift. The spatial resolution has been changed so considerably that pixel size has become smaller than a size of car which was earlier bigger than two or three buildings. This led to research on new classification algorithms for high and very high resolution remote sensing images because traditional pixel based analysis was proved to be insufficient due to its incapability to handle the internal variability of complex scenes (Schiewe, 2002; Blaschke and Strobl, 2001; Carleer et al., 2005). These also propelled object based approach or Object Based Image Analysis (OBIA) for very high resolution image segmentation (Hay and Castilla, 2006). Detailed applications and discussion on the development trends of OBIA can be found in Blaschke (2010). However, in this paper applications based on OBIA are not the concern. This paper deals with technological aspect of image segmentation, which concern about identification of objects but not much related to further analysis of the object. Still object analysis is required for assessment of segmentation accuracy.

According to the aforementioned definition of segmentation, the major thrust is on determining the suitable homogeneity measure which can discriminate the objects from each other. Some examples of the measures may be spectral, shape, texture and contexture. Most of the methods applied on remote sensing imageries are imported from other fields (Color image segmentation, Medical Image segmentation etc) and they work well because the underlying principal is same. For example, Cheng et al. (2001) extended the application of monochrome (single band) segmentation method, which was originally used on medical imagery, to colour image segmentation (three bands).

With the numerous recent developments of new segmentation methodologies, the requirement of their categorisations based on successful applications have become essential. Therefore, the

* Corresponding author.

first objective of this paper is to categorise the technologies of image segmentation by conceptualising the implementation details. Image segmentation techniques which are applied on optical remote sensing image segmentation are included whereas those applied on active remote sensing satellite imagery like SAR imagery are excluded because of the reason already mentioned. However, in order to state the technological development some of the non-remote sensing applications are stated too. The second objective of this paper is to give an insight to the readers about the state of art of technological aspects of image segmentation and aid in deciding the mathematical form for image segmentation.

The rest of the paper is organized as follows. Section 2 discusses about the development of segmentation as per the existing review papers on image segmentation. Section 3 describes the categorisation of image segmentation from broad to fine level. Section 4 states the conclusion of the performed review. In order to state the development in a particular technology, similar methods are grouped and presented in a paragraph in rest of this literature.

2. DEVELOPMENT OF SEGMENTATION

One of the early application of image segmentation on remote sensing points to ECHO (Extraction and Classification of Homogeneous objects) classification by Kettig and Landgrebe (1976). This states that association of segmentation with remote sensing imagery was not much later than the operation of the first remote sensing satellite Landsat-1 in 1972. There have been many developments in remote sensing image processing techniques after that. Haralick and Shapiro (1985), Reed and Buf (1993), Spirkovska (1993) and Pal and Pal (1993) did extensive review on early stage of image segmentation techniques existed used in various applications along with remote sensing. Developments of image segmentation algorithms for remote sensing imageries have been drastically increased after the availability of high resolution imagery (Schiewe, 2002; Blaschke, 2010). This is obvious with the failure of pixel based techniques on high resolution imageries as discussed in the introduction section. Further, the commercially available software eCognition, since 2000 based on Fractal Net Evaluation Approach (FNEA), incorporating similarity of objects at hierarchical scale, has revolutionised the research on image segmentation and is still influencing the research very substantially (Baatz and Schäpe, 2000; Blaschke, 2010). This is why most of the review papers before the period of the year 2000 don't specifically cover remote sensing applications. After that we do have a few good review papers. For example, Schiewe (2002) categorised the available remote sensing technologies for high resolution imagery, Carleer et al. (2005) evaluated qualitatively some of the most widely used image segmentation technologies for very high spatial resolution satellite imagery, Shankar (2007) presented various techniques with mathematical details of image segmentation techniques and Blaschke (2010) on OBIA.

3. CATEGORISATION OF SEGMENTATION

The abundance of literature on image segmentation makes the categorisation both necessary and challenging. The approach of categorisation in this paper is supplementary to some earlier review papers mentioned in section 2. (Reed and Buf, 1993; Pal & Pal 1993; Spirkovska, 1993; Schiewe, 2002; Shankar, 2007). Most of the earlier literatures have categorised them as a) Edge based b) Point/Pixel based c) Region based and d) Hybrid approach. Guo et al. (2005) categorised them as colour based and texture based algorithms. However, a more clear delineation is required considering the techniques which are used to achieve segmented objects.

A more general method of categorisation based on approach towards image analysis and applicable even beyond image segmentation domain are the bottom-up and top-down approaches. In image segmentation domain, they are often stated as model driven (top-down) and image driven approach (bottom-up) (Guindon, 1997). In this paper, this approach is stated as first stage of categorisation. It can also be stated as segmentation control based categorisation. However, in eCognition/Definiens developer software top-down and bottom-up approach refers to hierarchy of segmentation (eCognition Elements User Guide, 2004). It can be said that bottom-up approach forms object by combining/merging pixels or group of pixels whereas top-down approach moves from splitting the whole image into image objects based on heterogeneity criteria (Benz et al., 2004). However, this is not the only definition.

The second stage of categorisation points to features or homogeneity measures based approaches used to delineate image objects. The third stage of categorisation is based on operations on image used to generate image objects. These are edge detection, region growing/splitting and may be both of them. It is important to note that these stages are highly interrelated and generally developed methods pick up one or more methods from the list at different stages to perform final segmentation. For example, Beveridge et al. (1989) used thresholding object/background model for generating initial regions and region merging algorithm with spectral, shape and connectivity as homogeneity measures. Tilton (1996) used both region growing and edge detection for Landsat TM data. A detailed description of the categorisations and their inter-relationships are stated in the subsequent sections. Apart from aforementioned categorisation, image segmentation can also have supervised and unsupervised approach. Unsupervised segmentation holds its proximity to feature extraction and clustering whereas supervised segmentation incorporates segmentation accuracy as an addition to unsupervised scheme.

3.1 Image Driven approach

Image driven approach operates directly on the image pixels and detects objects solely based on the image features (Maxwell, 2005). Image driven approach extracts object based on the statistical features of the image derived from the pixels. This includes most of the solely edge based segmentation techniques. Edge based techniques detects edges and then closes the regions by contour generating algorithms (Schiewe, 2002). Canny Deriche operator is considered as good edge detector for remote sensing purposes (Carleer et al. 2005). However, different algorithms can also be tried. For example, Chehdi et al. (1993) used zero crossing of second derivative along four major directions to detect edge points and consequent closing of edges to generate regions of SPOT image. Edge detection is now more used for feature extraction in remote sensing and in segmentation of medical imagery (Pham et al., 2000). However, watershed transform is the current edge based segmentation technique being utilised in segmentation (Carleer et al., 2005).

3.2 Model Driven approach

Model based approach assumes that objects in an image are present in a certain pattern. Interested readers can look into

Rosenfield and Davis (1979) for more understanding of image models and segmentation. A list of models generally used for image segmentation are a) Object Background/Threshold Model, b) Neural Model, c) Markov Random Field Model, d) Fuzzy Model, e) Fractal Model, f) Multi-resolution and g) Transformation model namely Watershed model and Wavelet model. MRF model, Fuzzy model, Fractal model and Neural model have been widely studied previously (Pal and Pal, 1993; Reed and Buf, 1993). Therefore, the newly developed model comprises Watershed model and multi-resolution model. Fractal model has not much significant application in remote sensing and wavelet model is inherited in multi-resolution model. Hence, except fractal model all the models and their developments, approaches and applicability are described in subsequent sub-sections.

3.2.1 Object-background Model: Object Background models are based on histogram thresholding. They are primitive models for image segmentation. They follow a concept that there is a uniform background and objects are irregularly placed on this background (Rosenfield and Davis, 1979). They are mainly based on spectral properties. Spectral variation is represented by image histogram. This makes image histogram the choice for object delineation. Hence, finding an appropriate threshold between object and background fulfils the task of object identification. Most of the threshold based method follows an image model. In the next paragraph, some of them are discussed.

The widely used bi-level thresholding techniques have underlying object and background modelling (Weszka, 1978). Threshold can also be calculated based on the maximisation of class (object and background) separability error/ discriminant analysis (Otsu, 1979), maximisation of entropy based on the assumed probability distribution model (Pal and Pal, 1991) and many more. A detailed review on thresholding techniques can be found in Sahoo et al. (1988). Fuzzy thresholding approaches are the current developments in this field (details in fuzzy model section).

Currently, thresholding based methods are not popular in remote sensing areas especially in urban remote sensing applications with high resolution imagery. This is because of high degree of variation of histogram and hidden clustering problem (Beveridge et al., 1989).

3.2.2 Markov Random Field Model: Markov random field (MRF) model is not so old in remote sensing applications as compared to histogram thresholding. MRF model was conceptualised from Ising model (pp.1-23, Kinderman and Snell, 1980). MRF model takes into account the neighbourhood relationship which makes it attractive for modelling texture and contexture of images. The detailed mathematics of types of MRF models and their estimations can be found in the book by Li (pp. 21-47, 2009). However, a short summary of applications on remote sensing image segmentation is presented here.

One of the seminal papers of MRF in segmentation is Hansen and Elliot (1982). In remote sensing, the application of MRF was much later by Jeon and Landgrebe (1992). They used MRF for contextual classification of Landsat TM temporal data (pp. 243, Richards and Jia, 2006). Bouman and Shapiro (1994) applied unsupervised segmentation scheme with modified MRF model and named the model as multi-scale random field model (MSRF). MSRF used hybrid structure of quadtree and pyramid graph for scale representation. Then, expectation maximisation (EM) algorithm used for solving sequential maximizing a posteriori (SMAP) whose solution calculates the required parameters of MSRF model. They used multispectral SPOT image for their experimental results. Spectral and spatial features were used in MSRF model. Raghu and Yegnarayana, (1996) used supervised scheme for segmentation. They applied Gabor filters, for extracting texture feature, constituting a multi-resolution feature extraction mechanism. Texture feature vector was represented as Gaussian distribution and a posteriori probability scheme was formulated for assigning a partition label to a pixel where partition is expressed as noncausal MRF. Posterior probability of segmentation model was represented as Gibbs distribution and maximizing a posterior probability was done using Hopfield neural network with a deterministic relaxation modelling. This process used spectral, texture, spatial and prior knowledge as prior distribution. Technique was tested on three band image of IRS satellite. This paper used spectral, spatial and texture, in form of local interactions and class information. Both of the above method has used multi-resolution concept (see section 3.2.5). Jung et al. (2005) used multi-resolution MRF based unsupervised texture segmentation using Discrete Wavelet Transform (DWT). A MRF model was applied on each sub-band image separately, obtained from DWT considering spatial adjacency relationship. Parameter estimation was done by least squares estimate of Pseudo-maximum Likelihood. MAP criterion was optimized using simulated annealing (SA). Landsat TM was used for generating results through Gaussian MRFs. The properties used are spectral, spatial, contextual (spatial adjacency rule and clique functions) and texture.

Tsai and Tseng (1997) developed unsupervised segmentation scheme in which RGB of SPOT satellite was transformed into HSI colour space to estimate the number of colour sets by scale space filter based histogram thresholding. Then, iterated conditional mode (ICM) algorithm was employed for MAP estimation of GMRF based pixel partition labelling. Method used spectral and spatial information using texture (hybrid of local and global texture information) features for pixel based segmentation. Tseng and Lai (1999) also used GMRF but approximation was done by using Genetic algorithm instead of ICM for MAP estimation.

Sarkar et al. (2000) developed a modified technique to reduce the complexity of MAP-MRF estimation. Instead of working directly on pixels, they used a two stage algorithm for over-segmented image. At first stage, region adjacency graph was plotted for those regions. Energy function of MRF model was defined based on intra-region homogeneity and inter-region dissimilarity. At second stage, region merging is performed based on these energy equations value compared with a threshold based on Fischer distribution. This is an unsupervised MRF model based region merging approach which utilised spectral, spatial and textural properties. Sarkar et al. (2002) extended the above mentioned MRF based unsupervised segmentation approach for multiband imagery and used it for land-use classification.

D'Elia et al. (2003) modified Tree structured MRF model based on binary split of the image regions at each step. Initially, the regions were split in a binary tree pattern based on splitting criterion. In order to reduce fragmentation, estimation of field parameters was locally adaptive and a region merging parameter was also included. They modelled the image as a linear combination of original value plus zero-mean Gaussian noise. Estimation of the field parameters were based on local neighbourhood characteristics using maximum pseudo-likelihood estimation. Finally, MRF labelling was performed

using MAP estimation through iterative conditional mode approximation technique. Poggi et al. (2005) used the Tree-structured MRF model for supervised texture segmentation on multi-spectral spot data which uses prior knowledge about the class and its estimated parameters.

In some other segmentation applications, Yang et al. (2008) used MRF in fusion based segmentation of SAR and Landsat imagery. They used region adjacency graph for MRF model and region reliability measure based on image properties for fusion. Moser and Serpico (2008) used graph based multiscale segmentation and fused the feature of those segmentations at coarse and fine scales to get final segmentation.

MRF models have attracted quite a decent amount of research for image segmentation. This is because of its ability to integrate spectral, textural, contextual, spatial properties of image and even prior knowledge in form of prior distribution. However, the mathematical formulation and high computational complexity are the drawbacks.

3.2.3 Fuzzy Model: Fuzzy theory had been conceptualised by Zadeh (1965). It has been applied in various fields of engineering applications. Fuzzy segmentation adds fuzzy boundary for objects. In the subsequent paragraphs, few developments and fuzzy logic based techniques applied in remote sensing image segmentation would be stated.

In early remote sensing, fuzzy segmentation was derived from clustering methodology. In order to be tuned with the terminologies of research papers, clusters and segments are used interchangeably in this literature. Cannon et al. (1986) utilised fuzzy c-means clustering for image segmentation. Fuzzy c-means clustering is a form of minimizing within group sum of squared (WGSS) error. Each pixel holds a membership value derived from local minimum of WGSS error. Two methods used for hard clustering was confusion matrix oriented merging (percentage of total pixels in that cluster) and minimal spanning tree merging whose nodes are cluster centres and edges are distance between cluster centre. Here, the class information was already available which helped in pruning the spanning tree to form segments. Krishnapuram and Kellel (1993) modified fuzzy c-means by possibilistic c-means. They introduced a scale parameter to modify the objective function of original fuzzy c-means. This method doesn't the need of stating the number of clusters beforehand and is robust even in the presence of noise and outliers. Hence, filtering step may be avoided. However, this method requires a reasonable scale parameter value and good initialization. Thus, restricts its capability of automated segmentation. Fan et al. (2009) proposed a single point iterative weighted fuzzy c-means which uses prior knowledge for initialising cluster centres and spatial and spectral information for weighing the original fuzzy c-means distance calculation.

Caillol et al. (1993) incorporated fuzzy sets in Gaussian Markov random field model to segment image. They introduced an interesting approach in the sense that their method incorporates both hard and fuzzy segmentation simultaneously. They named their method as fuzzy stochastic estimation maximization. However, their approach was limited to two class segmentation. They primarily used grey level values. Tzafestas and Raptis (2000) used an iterative fuzzy clustering which can incorporate image properties namely, spectral, spatial, texture and frequency in fuzzy form for segmentation. The algorithm applied is locally adaptive and number of output clusters/segments is not fixed a priori. Thus, it produces optimum number of segments till it reaches a predefined threshold.

Pal et al. (2000) used fuzzy techniques for histogram thresholding based segmentation. They used fuzzy entropy, fuzzy geometry, fuzzy correlation and fuzzy clustering techniques for thresholding. Results were demonstrated on IRS and SPOT satellite imagery. Bandyopadhyay (2005) used genetic algorithm for fuzzy clustering. He included spatial information by incrementing pixel vector with mean of a 3x3 or higher neighbourhood. Then, spatial information was included using up-down pixel value difference from centre pixel. Wuest and Zhang (2009) have modified the Hierarchical Split and merge algorithm (HSMR) to perform an unsupervised segmentation. They used fuzzy band ratio to describe regions by their class densities. Then, fuzzy logic was used for comparing the region similarity. The algorithm was applied on Quickbird imagery and segmentation is basically proposed for land use purposes.

Most of the fuzzy segmentation methods are derived from fuzzy c-means clustering and fuzzy thresholding (Shankar, 2007). However, it is possible to incorporate fuzzy model in the most of the existing segmentation model e.g. Fuzzy MRF stated here and Fuzzy Neural models to be stated in next section. The decision of incorporating fuzzy model is based on the achievable complexity level of the segmentation.

3.2.4 Neural Model: Neural networks are based on simulation of human brain processing element called as Neurons. The structure of a neuron is shown in the fig-1. Rectangular blocks correspond to input multiplied by weights (W_i) and F correspond to threshold function and z correspond to linear sum of weights multiplied with corresponding input. One can build a network by increasing the number of neurons and number of layers or outputs, adding elliptical blocks in horizontal and vertical fashion. Layers in between input and output layers are known as hidden layers. The basis of Neural network lies in training of neural network. The aim of training is to model the process of data generation such that it can predict the output for unforeseen data. Training is generally associated with supervised methodology. However, unsupervised network can also be formulated e.g. Adaptive Resonance Theory 1 (ART1), ART2, Fuzzy ART and Self-organizing Maps (pp. 102-147, Tso and Mather, 2001).

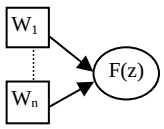


Figure 1. showing structure of a neuron

One of the early applications of neural networks in image segmentation is by Visa et al. (1991). They used co-occurrence matrix based texture feature vectors as input to self-organizing map (SOM) neural networks. Their aim was cloud detection from NOAA-10 and NOAA-11 satellite imagery. They used texture features derived from spectral values. Solaiman et al. (1994) proposed an edge based segmentation by automatically tuning parameters of Canny-Deriche recursive filtering using a multi-layer perceptron (MLP) network. They utilised spectral and spatial properties. Chen et al. (1996) modified the learning technique of MLP network by first removing any hidden layer, then selecting a polynomial basis function as the activation function. Essentially the network was linearized by this modification. This linearization made it capable of being trained by a Kalman filtering technique. This reduced the training time compared to back-propagation training. This is essentially a

supervised technique because of MLP network. The process is based on intensity values or spectral properties.

Baraldi and Parmiggiani (1995) modified ART neural network to simplified ART neural network (SARTNN) such that it estimates much lesser user defined parameters than ART and also has capability to accept multi-valued input. The architecture is similar to ART and used Vector degree of Match (VDM) for comparison of multi-valued vector. This process doesn't require a priori number of processing elements. They applied the algorithm for clustering the Landsat-5 imagery and found better clustering. Chen et al. (1997) proposed a neuro-fuzzy scheme for image segmentation. In the first step, they transformed image using histogram based non-uniform coarse coding technique. This resolved the proportion ambiguity, observed with patterns having proportional relationships when used as input for ART, due to normalizing inputs. In the second step, ART2 was applied as neural network due to its unsupervised nature. This clustered the input pattern into the desirable number of classes. However, the final informative classes are still to be formed. This final step was performed using fuzzy clustering. This method used spectral and spatial information in the form that the probability that adjacent pixels belong to same class is large. This method was applied on three multispectral channels, Green, Red and Infrared of SPOT HRV sensor.

Kuntimad and Ranganath (1999) used Pulse coupled neural network (PCNN) for image segmentation. The essential feature of PCNN lies in its one to one correspondence to image pixels. Further, it requires no training and directly produces segmented objects with edges. PCNN has also capability to utilise the neighbourhood relationship. Li et al. (2007) used improved pulse coupled network for image segmentation of IKONOS imagery. Their modifications differ in the sense of linking of neurons, edge-preserved prior smoothing instead of just smoothing and reduce algorithm complexity.

ANN might not have caught the eyes of researchers of remote sensing image segmentation but it has wide applications in medical imagery with different type of ANN like SOM, MLP and Hopfield Network to name a few (Peterson et al., 2002). This may be due to the challenge of generalization in conventional neural networks (Atkinson and Tatnall, 1997). However, PCNN seems to be a promising approach for unsupervised image segmentation with its capability to incorporate neighbourhood relationship. This area needs further research.

3.2.5 Multi-resolution Model: Woodcock and Strahler (1987) stated the importance of factor of scale in identification of objects present in the remote sensing imagery. Scale of an object is stated as level of aggregation and abstraction at which an object can be described (Benz et al, 2004). An object which is smaller than the spatial resolution of image cannot be identified. It is because of inappropriate scale of object. Based on this, two problems can be stated. Firstly, if object size is large then high spatial resolution satellite would fragment the object and secondly, if object size is small then low spatial resolution may not even recognize it. With the availability of high resolution satellite imagery second problem has been eliminated. Now, the first problem is to be solved. A general idea of multiscale/multiresolution approach arose for solving this kind of problem. Multi-scale segmentation can go both ways from coarse to fine (top-down) and fine to coarse (bottom-up) (Zhong et al., 2005). The idea for coarse to fine level states that initial segmentation can be performed at coarse level and this initial segmentation acts as input to next finer level segmentation. The reverse is true for fine to coarse approach. However, in both approaches a threshold is defined to decide merging or splitting. This creates a hierarchical segmentation. Now the next concern is how to represent the multi-resolution/multiscale/hierarchical and segment based on this representation scheme. This is the topic of discussion in the next paragraph.

Bongiovanni et al. (1993) used pyramidal structure for multi-resolution segmentation. They assumed image to be bi-modal and based on spectral property a threshold is determined to assign bimodality. The method iteratively finds the bi-modality and then follows top-down approach to segment based on their bimodality. The representation scheme for this operation was pyramidal where each pyramid node had four children. Bouman and Shapiro (1994) applied multiscale representation using a hybrid of pyramid graph (at finer level) and quadtree (at coarse level) for a MRF based image segmentation (already described in sec. 3.2.2).

Baatz and Schäpe (2000) brought significant development in the research of multi-resolution segmentation for remote sensing imagery after the introduction of Multi-resolution/hierarchical segmentation using Fractal Net Evaluation approach (FNEA). FNEA represents the notion of hierarchy as fractal net because of the self similarity notion of fractals. Each coarser level gets the input from finer level and if an object is identified at coarser level then it repeats its representation at each finer level which is referred as similarity. The process starts with each pixel as objects and then subsequently merged based on the criteria in which merged region shouldn't exceed a defined heterogeneity threshold. This approach has capability to incorporate spectral, texture, spatial, shape, size, prior knowledge and contextual properties of image. This approach is incorporated in eCognition/Definiens Developer, a commercial software product. This software revolutionised the field of remote sensing image segmentation with its immense possibility to provide GIS ready information (Blaschke, 2010; Benz et al., 2004). The selection of parameters, scale, smoothness and compactness, for multi-resolution segmentation requires an expert knowledge which makes it semi-automatic. However, Maxwell and Zhang (2005) proposed a fuzzy approach which automatically selects the parameter of the segmentation used in multi-resolution approach.

Apart from the development of eCognition/Definiens Developer some other techniques are also developed. Chen et al. (2003) applied a top-down strategy for multiscale segmentation applied on SPOT HRV image. They performed discrete wavelet transform on first principal component, obtained from PCT of original bands, to obtain coarse scale image and applied clustering for coarse scale segmentation. Segmentation at fine scale used region growing procedure based on seed pixel of region. Pixels are grouped with seed pixel based on spectral and textural feature vector. Grouping is based on a threshold of acceptable heterogeneity after merging. This process is performed iteratively until all coarse scale segmentation was subjected to fine scale segmentation. Zhong et al. (2005) used a bottom–up approach for multi-scale segmentation on IKONOS image using. Starting with a pixel region, region is subsequently merged with other based on achieved homogeneity measure. Homogeneity measure is derived from colour feature, smoothness feature and compactness feature. Once, every region at a particular scale are processed like this, the average size of objects are calculated. If average size satisfies the size

threshold then segmentation is optimal otherwise segmentation is carried on with the achieved regions.

Li et al. (2008) applied multiscale segmentation using hybrid of statistical region merging (SRM) for initial segmentation and minimum heterogeneity rule (MHR) for merging objects for high resolution Quickbird imagery. SRM utilises spectral, scale and shape measures for initial segmentation. Segmentation using SRM follows region growing technique where region growing is based on statistical test. Minimum Heterogeneity rule used colour (spectral) and shape property for region merging.

Multi-resolution model is indeed the most sought after technique for remote sensing image segmentation. It is possible to combine the concept of multi-scale to any other segmentation approach e.g. with MRF model (Bouman and Shapiro, 1994). Its combination with watershed model will be defined in next sub-section. Its success lies in its capability to incorporate spectral, shape, size, texture and contexture features of region at various scales for efficient segmentation especially for high resolution complex landscape imageries. The most typical part of this model is appropriate scale representation and information extraction from them (Chen et al., 2009). The method developed by Chen et al. (2009) aids in identifying the scale of proper representation of objects.

3.2.6 Watershed Model: Watershed model is a mathematical morphological approach and derives its analogy from a real life flood situation (Beucher, 1992). It transforms image into a gradient image. Then, image is seen as a topographical surface where grey values are deemed as elevation of the surface at that location. Then, flooding process starts in which water effuses out of the minimum grey value. When flooding across two minimum converges then a dam is built to identify the boundary across them. This method is essentially an edge based technique (Carleer et al., 2005). The original watershed algorithm was susceptible to over-segmentation so a modified marker-controlled based watershed algorithm was proposed by Beucher (1992). Watershed algorithm produces over-segmentation because of noise or textured patterns. The application of watershed algorithm on remote sensing imageries is relatively recent than other models. Next few paragraphs describe several modifications on marker-controlled watershed algorithm to reduce over-segmentation problem.

Traditionally watershed algorithm was applied with median filter to eliminate noise and preserve contours (Carleer et al., 2005; Sun and He, 2008). Chen et al. (2006) stated that median filter fails to encounter high imagery texture, generally present in high resolution imagery. They proposed a modified technique to encounter this problem. They used a non-linear filter named Peer group filtering for removal of noise and image smoothing. Then, a floating point based rainfall algorithm for watershed transformation was applied for initial segmentation. Then, a multi-scale region merging algorithm was applied based on spectral, shape and compactness feature for final segmentation. The algorithm was applied on IKONOS imagery. Chen et al. (2008) proposed a different gradient operator for watershed transform which efficiently reflect texture information. The gradient image used is known as Homogeneity gradient image or H-image. H-values are calculated by a local window based operation. Dark and bright areas in H-image represent region centers and region boundaries. A rainfalling algorithm for watershed transformation was used followed by region merging, where regions were represented using Region adjacency graph. Region merging was based on colour, texture and shape features. Algorithm was applied on SPOT three band image with 2.5 m resolution.

Watershed algorithm is new segmentation approach with relatively less application in remote sensing image segmentation than other described models. However, it may be good for initial segmentation in a multi-scale resolution as it produces an over-segmentation. Over-segmentation elimination is also a problem associated with this method which needs further research. The commercial software ERDAS Imagine Extension (IMAGINE WS) has incorporated this algorithm.

3.3 Categorisation based on homogeneity measure

Next stage of categorization corresponds to the homogeneity measures used for image segmentation. But before that it is necessary to determine the possible homogeneity measures of image features. This requires a well understanding of image objects and the final outcome of image segmentation. Image objects are real world objects represented on remote sensing image. With very high resolution satellite, image objects can be visualized by human eye. This has been addressed by some researchers. For example, Wang and Terman (1997) suggested sensory cues of segregation based on Gestalt psychology for segmentation and Fu and Mui (1981) as psycho physical perception problem for segmentation. It is similar to elements of analysis for image interpretation by human eye (pp. 67-68, Richards and Jia, 2006). Thus, the possible measures are based on similarity comprises spectral, texture, spatial, size, shape, and temporal. Some other semantic information prior knowledge, context and connectedness are also required (Wang and Terman, 1997).

The primary homogeneity measure is spectral/tonal feature. Secondary homogeneity measures are spatial, texture, shape and size. Tertiary homogeneity measures are contextual, temporal and prior knowledge (pp. 67-68, Richards and Jia, 2006). As per the order, the most important is primary then secondary and then tertiary. Secondary and tertiary measures are more important when the boundaries of objects are required to be precise with very less mis-segmented pixels. In this study, more emphasis is given on secondary and tertiary measures which were not widely covered in earlier literatures. The list of measure may not be exhaustive but surely cover most of the available techniques existing for image segmentation. Subsequent sections describe the trend of techniques for different homogeneity measures used in image segmentation.

3.3.1 Spectral and Textural Features: The most primitive measures of homogeneity are spectral and textural features. Spectral values refer to grey levels or pixel values of an image. It has been long realised that using only spectral features good segmentation results are not possible but was still practiced due to the ease of incorporating them in digital format (Kettig and Landgrebe, 1976). Texture features points to spatial pattern represented by spectral values (Haralick et al., 1973). A textured image may have various texture patterns. However, quantitatively characterizing texture is not simple (pp. 128-130, Richards and Jia, 2006). Due to this fact texture segmentation has been studied widely generally in combination with other features like shape, spectral and contextual and various models till today.

Chen and Pavlidis (1978) used co-occurrence matrix and a quadtree based structure to determine texture similarity for grouping pixels in a region. Cross et al. (1988) also used quadtree based hierarchical structure and applied texture

measure was local difference statistics. Guo et al. (2005) used texture measure derived from local binary pattern and used wavelet transform to pre-process the image and derived texture from local binary pattern. They also used quadtree structure for splitting and merging. It can be seen from trend that quadtree based hierarchical image splitting has been the trusted method of texture segmentation for decades.

Conners et al. (1984) used spatial grey level dependence method (SGLDM) and six texture measures namely inertia, cluster shade, cluster prominence, local homogeneity, energy and entropy in region growing algorithm based on split and merge tecnique. Ramstein and Raffy (1989) used variogram and fractal dimension measures for texture segmentation and classification. Ryherd and Woodcock (1996) used a 3x3 adaptive window to calculate texture image based on local variance and applied a multi-pass region growing algorithm which builds spatial homogeneous objects using Euclidean distance in n-dimensional space. They showed that segmentation accuracy of derived texture image is better when compared with original image, used spectral property only. Algorthm was tested with SPOT panchromatic image.

Texture segmentation is one of the most sought after segmentation technique. It is evident from Reed and Buf (1993) and the above literature. This is mainly because of the presence of highly textured regions in high resolution satellite imagery. Currently, the research has shifted from texture to multi-resolution model.

3.3.2 Shape and Size Features: The importance of shape and size measure could be understood when the natural object are to be identified on satellite imagery. For example, a river and a pond may has same spectral, texture and spatial properties but they differ in shape and size. It is because rivers are linear and unbounded features whereas ponds are non-linear and bounded features. Shape and size measures are mostly utilised as complementary to each other. Further, they are always applied in conjunction with the spectral and texture measures. Only some substantial algorithms based on the recent developments are mentioned.

Beveridge et al. (1989), performed over-segmentation and then utilised shape, connectivity and size measure for region merging to achieve segmentation. Multi-resolution models represent the size of object through spatial scales (Bongiavanni et al., 1993). Fractal Net Evaluation approach (explained in section 3.2.5) used in commercial software, eCognition/Definiens developer, also uses scale, shape and compactness parameter.

The state of art use of shape and size refers to multi-scale/multi-resolution approach to image segmentation. Shape and size measures are especially helpful when delineating complex objects in high resolution satellite imagery.

3.3.3 Context: Context generally refers to spatial context which means relationship of pixels with its neighbourhood (Thakur and Dikshit, 1997). Contextual information is also used in conjugation with spectral or texture or both measures. Few methods are found which utilise specifically context based segmentation. Context helps in avoiding fragmentation of a segment and merging. For example in an urban image, cars in a parking lot may cause fragmentation unless context measures are applied.

A good recent example of context based segmentation is Fan and Xia (2001). They deduced context information from spatial and scale space of image and modeled five context models with quadtree model for scale dependency. They called their algorithm as multi-contextual (due to five context models) and multi-scale approach to Bayesian segmentation which in mathematical terms solves context-based mixture model likelihood. They used their methods for aerial and SAR imagery. Even eCognition/Definiens Developer software has the capability of including the context information based on neighbourhood relationship measures. Benz et al. (2004) demonstrates in the paper that how eCognition integrates spatial and scale context as semantic information in identifying the appropriate image objects. Contextual constraints are used in segmentation and classification and are well modeled by Markov Random Field. This is why several context-based classifications use MRF model (Melgani and Serpico, 2003; Jackson and Landgrebe, 2002).

Context is especially useful when segmentation requires bigger area to be identified as one segment e.g. land use classification. MRF models are currently the best model for implementation of contextual measures.

3.3.4 Temporal: Temporal measure refers to measurement based on images of same area and sensor characteristics in different time (pp. 67-68, Richards and Jia, 2006). Temporal measure is not directly used for segmentation but is used as an application of segmented image.

Carlotto (1997) performed temporal segmentation for change detection from Landsat TM. He used total difference image to segment based on histogram thresholding. Jeansoulin et al. (1981) performed segmentation using fuzzy edge detection and region growing for segmentation and demonstrated how temporal criterion can be used to detect changes based on objects. Hanaizumi et al. (1991) used spatial segmentation for change detection and showed result on Landsat TM imagery. They used division and detection procedure where divison/region-splitting was performed by fitting regression model on pixel scattergram. Dambra et al. (1991) fused multi-temporal imagery using segmented image. SAR segmented image is also used for change detection. Several SAR segmenting methods are reviewed by Caves et al. (1996).

Yamamoto et al. (2001) detected change in SPOT HRV and Landsat TM image using 3-D segmentation with time as Z axis. They applied local statistical regression model for region splitting using spatial and spectral measures. Hall and Hay (2003) used multi-object scale analysis for change detection which utilises Marker Controlled watershed segmentation (Beucher, 1992). Lhermitte et al. (2008) introduced multi-temporal hierarchical image segmentation. They segmented the 10 daily data of SPOT VGT sensor by first decomposing original image time series in Fast Fourier Transform component and then performed hierarchical segmentation analogous to eCognition (Baatz and Schäpe, 2000) using Euclidean distance between FFT components of same frequency as measure of similarity.

Temporal characteristics have important application in monitoring changes like land-use change, disaster mapping, traffic flows, crop mapping etc (pp-280-81, Campbell, 2007). Temporal segmentation has been used mainly for change detection in a series of temporal image. Its application is mainly seen for large area change detection rather than small area. Thus, more applications have been found on low resolution images than high resolution.

3.3.5 Prior Knowledge: Prior knowledge refers to photointerpreter knowledge regarding the regions/objects of the image (pp. 342-352, Richards and Jia, 2006). It may be the knowledge of classes of the image region or about some specific area, building or trends etc. Incorporating prior knowledge in image analysis is one steps towards developing artificial intelligence in the machine (Srinivasan and Richards, 1993). Prior knowledge may not be the primary measure for segmentation but it has the capability of utilising the location based information. For example, it is our prior knowledge which generally says that small buildings mean residential areas and large buildings means commercial or institutional areas. This indicates towards differentiation based on shape properties. In the next paragraph, few prior knowledge based segmentation or prior knowledge based homogeneity measure derivation are described.

Ton et al. (1991) divided segmentation techniques into two types as partial segmentation (without using a priori knowledge) and complete segmentation (using a priori knowledge). The approach for knowledge based can be further divided into histogram-oriented and cluster-oriented (Ton et al., 1991; Paudyal et al., 1994). Most of the popular method like Hierarchical split and Merge (Ojala and Pietikainen, 1999), region growing, multi-resolution used by eCognition (Baatz and Schäpe, 2000) etc are partial segmentation techniques. Ton et al. (1991) used spectral and spatial knowledge rules for supervised segmentation of Landsat TM image. They automated generation of spectral knowledge based rules based on training data and hierarchical classification. They applied both threshold and region growing for segmentation.

Liu et al. (1993) used texture measure for region uniformity and contexture information at pixel level for segmentation. They used knowledge in determining the best texture measure, which gives minimum error using multivariate Gaussian Bayesian classifier, out of the available for good segmentation. The method used is essentially supervised segmentation. Using similar concepts some researchers incorporated knowledge in textural measures (Paudyal et al., 1994; Simman, 1997).

Smits and Annoni (1999) used no prior information but derived knowledge, automatically from a selected region, to select the best feature which can distinguish object from its neighbours. Jinghui et al. (2004) also used GIS prior information to extract building from Quickbird imagery using fuzzy connectedness algorithm.

Poggi et al. (2005) used tree structured MRF model in incorporating prior knowledge for supervised segmentation. Benz et al. (2004) also showed how expert knowledge can be included in segmentation based fuzzy classification.

Prior knowledge is incorporated in mathematical models by using class distribution information. In fuzzy models, it can be incorporated as semantic rules (Benz et al., 2004). Prior knowledge is specifically useful when for segmentation of complex landscape object indistinguishable using texture and context.

4. CONCLUSIONS

With the numerous amounts of image segmentation techniques presented in this paper, it might be possible to get confused regarding what is presented in this paper. Thus, it is important to summarize all of those to regain the content of this paper.

Image segmentation methodologies were categorized in three stages. At first stage comes model driven approach and image driven approach (mainly based on statistical analysis). The second stage corresponds to homogeneity based measure, and final category corresponds to mode of operations on an image, e.g. edge detection, region growing/splitting.

In model driven approach, object background model is insufficient for remotely sensed imagery. Neural model generally suffers from complexity regarding decision of network structure, proper learning and generalization of network. Hence, neural model is not one of the liked approaches by most of the researchers. Markov Random Field model has attracted quite a decent research in image segmentation. It can utilise significant image properties namely, spectral, spatial, texture, contexture and prior knowledge. However, MRF lacks the integration of shape and size and implementation of MRF is very complex.

Fuzzy model has been applied in remote sensing image segmentation mostly by means of fuzzy clustering of image or fuzzy thresholding. The strength of fuzzy model lies in ambiguity resolution. It can easily ensemble itself with neural model, MRF model and also histogram thresholding (Chen et al., 1997; Caillol et al., 1993; Shankar, 2007).

Multi-resolution (MR)/Multi-scale model is the most widely used model in remote sensing image segmentation. It has also been incorporated in a commercial software eCognition/ Definiens Developer. This model is capable identify object and object features at its intrinsic scale which makes object extraction of various scales possible (Chen at al. 2009). The problem of MR approach is scale representation and information extraction from each scale. The idea of MR approach is complex but when appropriately implemented has wide usage especially in remote sensing satellite images dealing with urban areas.

Watershed model based on mathematical morphological operators is another budding technology with respect application in remote sensing image segmentation. Further, research on this approach is required.

Homogeneity measures described in this paper are spectral, spatial, texture, shape, size, contextual, temporal and prior knowledge. Spectral measure is the most primitive one and quite long it has been realised that this alone wouldn't be able to deal with high resolution satellite imagery (Zhong et al., 2005). The second most widely applied homogeneity measure is based on texture. Texture segmentation is more successful because it inherits spectral and spatial properties in itself. However, this would still not yield a perfect segmentation. A better segmentation would require a model or methodologies which utilise most of the above mentioned measures to calculate region homogeneity or heterogeneity threshold. Integration of prior knowledge and contextual information has seen quite a good research in segmentation.

The selection of segmentation approach depends on what quality of segmentation is required. Further, it also depends on what scale of information is required. Few examples, based on done literature review in this paper, would be stated now to illustrate the idea. For urban GIS applications objects at different scale are required. For landuse coarse scale segmentation is required whereas for land cover fine scale. Hence, multi-resolution model would be the best choice. For highly textured image MRF model might be the good choice.

Fuzzy model would be good choice to represent ambiguity of region boundaries. Neural model would be good choice no prior distribution can be assumed and not very high quality object information is required. Among homogeneity measures, spectral, shape, size, scale, compactness and texture should be concerned when complex landscapes are to be analyzed.

As a part of future recommendation, some of the mentioned approaches in this paper should be implemented to look how each behaves on same image. Behaviours with images of different spatial resolution would be quite interesting. Further, addition of existing quantitative analysis of recent segmentation evaluation techniques would be quite helpful.

REFERENCES

Atkinson, P. M. and Tatnall, A. R. L., 1997. Introduction Neural networks in remote sensing. *International Journal of Remote sensing*, 18(4), pp. 699-709.

Baatz, M. and Schäpe, A., 2000. Multiresolution segmentation: An optimization approach for high quality multi-scale image segmentation. In: *J.Strobl, & T. Blaschke (Eds.), Angewandte geographische informationsverarbeitung*, Heidelberg: Wichmann-Verlag, Vol. XII, pp. 12-23.

Bandyopadhyay, S., 2005. Satellite image classification using genetically guided fuzzy clustering with spatial information. *International Journal of Remote sensing*, 26(3), pp. 579-593.

Baraldi, A. and Parmiggiani, F., 1995. A neural network for unsupervised categorization of multivalued input patterns: An application to satellite image clustering. *IEEE Transactions on Geoscience and Remote Sensing*, 33(2), pp. 305-316.

Benz, U. C., Hofmann, P., Willhauck, G., Lingenfelder, I. and Heynen, M., 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry & Remote Sensing*, 58(3-4) pp. 239–258.

Beucher S., 1992. The watershed transformation applied to image segmentation. *Scanning Microscopy Supplemen*t, 6, pp. 299–314.

Beveridge, J. R., Griffith, J., Kohler, R. R., Hanson, A. R. and Riseman, E. M., 1989. Segmenting images using localized histograms and region merging. *International Journal of Computer Vision*, 2(3), pp. 311-347.

Blaschke, T., 2010. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1), pp. 2-16.

Blaschke, T. and Strobl, J., 2001. What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. *GIS-Zeitschrift für Geoinformationssysteme*, 14(6), pp. 12-17.

Bongiovanni, G., Cinque, L., Levialdi, S. and Rosenfeld, A., 1993. Image segmentation by a multiresolution approach. *Pattern Recognition*, 26(12), pp. 1845-1854.

Bouman, C. A. and Shapiro, M., 1994. A multiscale random field model for Bayesian image segmentation. *IEEE Transactions on Image Processing*, 3(2), pp. 162-177.

Caillol, H., Hillion, A. and Pieczynski, W., 1993. Fuzzy random fields and unsupervised image segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 31(4), pp. 801-810.

Campbell, J. B., 2007. *Introduction to Remote Sensing.* Taylor and Francis, London, pp. 280-281.

Cannon, R. L., Dave, J. V., Bezdek, J. C. and Trivedi, M. M., 1986. Segmentation of a thematic mapper image using the fuzzy c-means clusterng algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, 24(3), pp. 400-408.

Carleer, A. P., Debeir, O. and Wolff, E., 2005. Assessment of very high spatial resolution satellite image segmentations. *Photogrammetric Engineering & Remote Sensing*, 71(11), pp. 1285–1294.

Carlotto, M. J., 1997. Detection and analysis of change in remotely sensed imagery with application to wide area surveillance. *IEEE Transactions on Image Processing*, 6(1), pp. 189 – 202.

Caves, R. G., McConnell, I., Cook, R. and Quegan, S., 1996. Multichannel SAR segmentation: algorithms and applications. *IEE Colloquium on image processing for remote sensing*, 27, pp. 2/1-2/6.

Chehdi, K. and Liao, Q., 1993. Satellite image segmentation using edge-region cooperation. In: *Proceedings of IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*, 1, pp. 47-50.

Chen, J., Pan, D. and Maz, Z., 2009. Image-object detectable in multiscale analysis on high-resolution remotely sensed imagery. *International Journal of Remote sensing*, 30(14), pp. 3585–3602.

Chen, K. S., Tsay, D. H., Huang, W. P. and Tzeng, Y. C., 1996. Remote sensing image segmentation using a Kalman filter-trained neural network. *International Journal of Imaging Systems and Technology*, 7(2), pp. 141-148.

Chen, P. C. and Pavlidis, T., 1978. Segmentation by texture using a co-occurrence matrix and a split-and-merge algorithm. In: *Proceedings of the 4th International Joint Conference on Pattern Recognition*, Kyoto, Japan, pp. 95-105.

Chen, Q. X., Luo, J. C., Zhou, C. H. and Pei, T., 2003. A hybrid multi-scale segmentation approach for remotely sensed imagery. In: *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 6, pp. 3416-3419.

Chen, S., Luo, J., Shen, Z., Hu, X. and Gao, L., 2008. Segmentation of multi-spectral satellite images based on watershed algorithm. In: *Proceedings of International Symposium on Knowledge Acquisition and Modeling, pp* 684-688.

Chen, S. W., Chen, C.F., Chen, M. S., Cheng, S., Fang, C.Y. and Chang, K.E., 1997. Neural-fuzzy classification for segmentation of remotely sensed images. *IEEE Transactions on Signal Processing*, 45(11), pp. 2639-2654.

Chen, Z., Zhao, Z., Gong, P. and Zeng, B., (2006) A new process for the segmentation of high resolution remote sensing imagery. *International Journal of Remote Sensing*, 27(22) , pp. 4991-5001.

Cheng, H.D., Jiang, X. H., Sun, Y. and Wang, J., 2001. Color image segmentation: advances and prospects. *Pattern Recognition*, 34(12), pp., 2259-2281.

Conners, R. W., Trivedi, M. M. and Harlow, C. A., 1984. Segmentation of a high-resolution urban scene using texture operators. *Computer Vision, Graphics and Image Processing*, 25(3), pp. 273-310.

Cross, A. M., Mason, D. C. and Dury, S. J., 1988. Segmentation of remotely-sensed images by a split-and-merge process. *International Journal of Remote Sensing*, 9(8), pp. 1329-1345.

D'Elia, C., Poggi, G. and Scarpa, G., 2003. A tree-structured Markov random field model for Bayesian image segmentation. *IEEE Transactions on Image processing*, 12(10), pp. 1259-1273.

Dambra, C., Rolit, F., Serpicot, S. B., Wielogorskiz, A. and Agostinelli, A., 1991. Remote sensing data fusion by means of a region-overlapping technique. In: *Proceedings of* IEEE *International Geoscience and Remote Sensing Symposium*, 2, pp. 1091-1094.

eCognition Elements User Guide, 2004 @copright Definiens Imaging, Germany. http://www.gis.unbc.ca/help/software /ecognition4/ELuserguide.pdf (accessed 20 Sep. 2009)

Fan, G. and Xia, X. G., 2001. A joint multicontext and multiscale approach to Bayesian image segmentation. *IEEE Transactions on Geoscience and Remote Sensing* 2001, 39(12), pp. 2680-2688.

Fan, J., Han, M. and Wang, J., 2009. Single point iterative weighted fuzzy c-means clustering algorithm for remote sensing image segmentation. *Pattern Recognition*, 42(11), pp. 2527—2540.

Fu, K. S. and Mui, J. K., 1981. A survey on image segmentation. *Pattern Recognition*, 13(1), pp. 3-16.

Guindon, B., 1997. Computer-based aerial image understanding: a review and assessment of its application to planimetric information extraction from very high resolution satellite images. *Canadian Journal of Remote Sensing*, 23(1), pp. 38-47.

Guo, D., Atluri, V. and Adam, N., 2005. Texture-based remote-sensing image segmentation. In: *Proceedings of IEEE International Conference on Multimedia and Expo*, pp. 1472-1475.

Hall, O. and Hay, G. J., 2003. A multiscale object-specific approach to digital change detection. *International Journal of Applied Earth Observation and Geoinformation*, 4 (4), pp. 311-327.

Hanaizumi, H., Okumura, H. and Fujiniura, S., 1991. Change detection from remotely sensed multi-temporal images using spatial segmentation. In: *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, 2, pp. 1079 – 1082.

Hansen, F. R. and Elliott, H., (1982). Image segmentation using simple Markov random field models. *Computer Graphics and Image Processing*, 20(2), pp. 101-132.

Haralick, R. M. and Shapiro, L. G., 1985. Image segmentation techniques. *Computer Vision, Graphics and Image Processing*, 29(1), pp. 100-132.

Haralick, R. M., Shanmugam, K. and Dinstein, I., 1973. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3(6), pp. 610-621.

Hay, G. J. and Castilla, G., 2006. Object-Based Image Analysis: Strengths, Weaknesses, Opportunities and Threats. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Tokyo, Japan, Vol. XXXVI, part 6.

Jackson, Q. and Landgrebe, D., 2002. Adaptive Bayesian contextual classification-based on Markov random fields. *IEEE Transactions on Geoscience and Remote Sensing*, 40(11), pp. 2454-2463.

Jeansoulin, R., Fontaine, Y. and Werner F., 1981. Multitemporal segmentation by mean of fuzzy sets. In: *Procceding* of *Seventh International Symposium on Machine processing of Remotely Sensed Data*, LARS Purdue University, Layfette, pp. 336-339.

Jeon, B. and Landgrebe, D. A., 1992. Classification with spatio-temporal interpixel class dependency contexts. *IEEE Transactions on Geoscience and Remote Sensing*, 30(4), pp. 663-672.

Jinghui, D., Veronique, P. and Hanqing, L., 2004. Building extraction in urban areas from satellite images using GIS data as prior information. In: *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 7, pp. 4762-4764.

Jung, M., Yun, E. and Kim. C., 2005. Multiresolution approach for texture segmentation using MRF models. In: *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, 6, pp. 3971-3974.

Kerfoot, I. B. and Bresler Y., 1999. Theoretical analysis of multispectral image segmentation criteria. *IEEE Transactions on Image Processing*, 8(6), pp. 798-820.

Kettig, R. L. and Landgrebe, D. A., 1976. Classification of multispectral image data by extraction and classification of homogeneous objects. *IEEE Transactions on Geoscience Electronics*, 14(1), pp. 19-26.

Kinderman R. and Snell J. L., 1980. Contemporary Mathematics: Markov random field and their applications, American Mathematical Society, Rhode Island, USA. http://www.cmap.polytechnique.fr/~rama/ehess/mrfbook.pdf (accessed 24 Aug. 2009)

Krishnapuram, R. and Keller, J. M., 1993. The possibilistic c-means algorithm: insights and recommendations. *IEEE Transactions on Fuzzy Systems*, 4(3), pp. 385-393.

Kuntimad, G. and Ranganath, H. S., 1999. Perfect image segmentation using pulse coupled neural networks. *IEEE Transactions on Neural Networks*, 10(3), pp. 591-598.

Lhermitte, S., Verbesselt, J., Jorckheere, I., Nackaerts, K., Aardt, J. A. N. V., Verstraeten, W. W. and Coppin, P., 2008. Hierarchical image segmentation based on similarity of NDVI time series. *Remote Sensing of Environment*, 112(2), pp. 506-521.

Li, H.T., Gu, H.Y., Han, Y.S. and Yang, J.H., 2008. An efficient multi-scale segmentation for high-resolution remote sensing imagery based on statistical region merging and minimum heterogeneity rule. *International Workshop on Earth Observation and Remote Sensing Applications*, 1, pp. 1-6.

Li, L., Ma, J. and Wen, Q., 2007. Parallel fine spatial resolution satellite sensor image segmentation based on an improved pulse-coupled neural network. *International Journal of Remote sensing*, 28(18), pp. 4191–4198.

Li, S. Z., 2009. *Markov Random Field modelling in Image Analysis*. Springer-Verlag, Heidelberg , New York, pp. 21-47.

Liu, S., Berthod, M. and Giraudon, G., 1993. Satellite image segmentation using textural information, contextual information

and map knowledge. In: *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, 3, pp. 355-360.

Maxwell, T., 2005. Object-oriented classification: Classification of pan-sharpening quickbird imagery and a fuzzy approach to improving image segmentation efficiency. MScE Thesis, Department of Geodesy and Geomatics Engineering Technical Report No. 233, University of New Brunswick, Fredericton, Canada, pp. 157.

Maxwell, T. and Zhang, Y., 2005). A fuzzy logic approach to optimization of segmentation of object-oriented classification. In: *Proceedings of SPIE 50th Annual Meeting - Optics & Photonics* San Diego, California, USA, 5909, pp. 1-11.

Melgani, F. and Serpico, S. B., 2003. A Markov random field approach to spatio-temporal contextual image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 41(11), pp. 2478-2487.

Moser, G. and Serpico, S. B., 2008. Classification of high resolution images based on MRF fusion and multiscale segmentation. In: *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2, pp. 277-280.

Ojala, T. and Pietikainen, M., 1999. Unsupervised texture segmentation using feature distributions. *Pattern Recognition*, 32(3), pp. 477-486.

Otsu, N., 1979. A threshold selection method from gray-level histograms. *IEEE Transactions on System, Man and Cybernatics*, 9(1), pp. 62-66.

Pal, N. R. and Pal, S. K., 1993. A review on image segmentation techniques. *Pattern Recognition*, 26(9), pp. 1274-1294.

Pal, N. R. and Pal, S. K., 1991. Entropy: a new definition and its applications. *IEEE Transactions on System man and Cybernetics*, 21(5), pp. 1260-1270.

Pal, S. K., Ghosh, A. and Shankar, B. U., 2000. Segmentation of remotely sensed images with fuzzy thresholding, and quantitative evaluation. *International Journal of Remote sensing*, 21(11), pp. 2269–2300.

Paudyal, D. R., Eiumnoh, A., Aschbacher, J. and Schumann, R., 1994. Knowledge based classification of multi-temporal ERS-1 and JERS-1 SAR images over the tropics. In: *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, 3, pp. 1612-1614.

Pavlidis, T., 1988. Image Analysis. *Annual review of Computer Science*, 3, pp. 121-146.

Petersen, M. E., De Ridder, D. and Handels, H., 2002. Image processing with neural networks - A review. *Pattern Recognition Letters*, 35(10), pp. 2279-2301.

Pham, D. L., Xu, C. and Prince, J. L., 2000. A survey of current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, 2, pp. 315–337.

Poggi, G., Scarpa, G. and Zerubia, J. B., 2005. Supervised segmentation of remote sensing images based on a tree-structured MRF model. *IEEE Transactions on Geoscience and Remote Sensing*, 43(8), pp. 1901-1911.

Raghu, P. P. and Yegnanarayana, B., 1996. Segmentation of Gabor-filtered textures using deterministic relaxation. *IEEE Transactions on Image Processing*, 5(12), pp. 1625-1636.

Ramstein, G. and Raffy, M., 1989. Analysis of the structure of radiometric remotely-sensed images *International Journal of Remote Sensing*, 10(6), pp. 1049-1073.

Reed, T. R. and Buf J. M. H. D., 1993. A review of recent texture segmentation and feature extraction techniques. *Computer Vision Graphics and Image Processing: Image Understanding*, 57(3), pp. 359-372.

Richards, J. A. and Jia, X., 2006. *Remote Sensing Digital Image Analysis: An Introduction*. Springer-Verlag, Heidelberg, New York, pp. 67-68; pp. 128-130; pp. 342-352; pp. 243.

Rosenfield, A. and Davis, L.S., 1979. Image segmentation and image model. Proceedings of IEEE, 67(5), pp. 764-772.

Ryherd, S. and Woodcock, C., 1996. Combining spectral and texture data in the segmentation of remotely sensed images. *Photogrammetric Engineering and Remote Sensing*, 62(2), pp. 181-194.

Sahoo, P. K., Soltani, S. and Wong, K. C., 1988. A survey of thresholding techniques. *Computer Vision, Graphics, and Image processing*, 41(2), pp. 233-260.

Sarkar, A., Biswas, M. K. and Sharma K. M. S., 2000. A simple unsupervised MRF model based image segmentation approach. *IEEE Transactions on Image processing*, 9(5), pp. 801-812.

Sarkar, A., Biswas, M. K., Kartikeyan, B., Kumar, V., Majumder, K. L. and Pal, D. K., 2002. A MRF model-based segmentation approach to classification for multispectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 40(5), pp. 1102-113.

Schiewe, J., 2002. Segmentation of high-resolution remotely sensed data-concepts, applications and problems. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Ottawa, Canada, Vol. XXXIV, Part 4, pp. 358-363.

Shankar, B. U., 2007. *Lecture Notes on Computer Science Series 4400; Transactions on Rough Sets VII* "Novel classification and segmentation techniques with application to remotely sensed images". Springer Berlin / Heidelberg, pp. 295-380.

Simaan, M. A., 1997. Techniques for large zone segmentation of seismic images. In: *Proccedings of IEEE International Geoscience and Remote Sensing Symposium*, 1, pp. 261-263.

Smits, P. C. and Annoni, A., 1999. Towards operational knowledge-based remote-sensing image analysis. *Pattern Recognition Letters*, 20(11-13), pp. 1415-1422.

Solaiman, B., Mouchot, M. C. and Koffi, R. K., 1994. Multispectral LANDSAT images segmentation using neural networks and multi-experts approach. In: *Proccedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 4, pp. 2109-2111.

Spirkovska, L., 1993. A summary of image segmentation techniques: NASA technical memorandum; 104022, California, USA. http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19940006802_1994006802.pdf (assessed 21 June, 2009)

Srinivasan, A. and Richards J. A., 1993. Analysis of GIS spatial data using knowledge-based methods. *International Journal of Geographic Information Systems*, 7(6), pp. 479–500.

Sun, Y. and He, G. J., 2008. Segmentation of high-resolution remote sensing image based on marker-based watershed algorithm. In: *Proceedings of Fifth International Conference on*

Fuzzy Systems and Knowledge Discovery, Jinan, Shandong, China, pp. 271-276.

Thakur, R.S. and Dikshit, O., 1997. Contextual classification with IRS LISS-II imagery. *ISPRS Journal of Photogrammetry & Remote Sensing*, 52(2). pp. 92-100.

Tilton, J. C., 1996. Hybrid image segmentation for earth remote sensing data analysis. In: *Proccedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*,1, pp. 703-705.

Ton, J., Sticklen, J. and Jain, A. K., 1991. Knowledge-based segmentation of Landsat images. *IEEE Transactions on Geoscience and Remote Sensing (IGARSS)*, 29(2), pp. 222–232.

Tsai, I. W. and Tseng D. C., 1997. Segmentation of multispectral remote-sensing images based on Markov random fields. In: *Proccedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 1, pp. 264-266.

Tseng D. C. and Lai, C. C., 1999. A genetic algorithm for MRF-based segmentation of multi-spectral textured images. *Pattern Recognition Letters*, 20(14), pp. 1499-1510.

Tso, B. and Mather, P. M., 2001. *Classification methods for Remotely Sensed Data.* Taylor and Francis, New York, pp. 102-147.

Tzafestas, S. G. and Raptis, S. N., 2000. Image segmentation via iterative fuzzy clustering based on local space-frequency multi-feature coherence criteria. *Journal of Intelligent and Robotic Systems*, 28(1-2), pp. 21-37.

Visa, A., Valkealahti, K. and Simula, O., 1991. Cloud detection based on texture segmentation by neural network methods. *IEEE International Joint Conference on Neural Networks*, 2, pp. 1001-1006.

Wang, D. and Terman, D., 1997. Image segmentation based on oscillatory correlation. *Neural Computation*, 9(4), pp. 805-836.

Weszka, J. S., 1978. A survey of threshold selection techniques. *Computer graphics and Image processing*, 7(2), pp. 259-265.

Woodcock, C. E. and Strahler, A. H., 1987. The factor of scale in remote sensing. *Remote Sensing of Environment*, 21(3), pp. 311-332.

Wuest, B. and Zhang, Y., 2009. Region based segmentation of QuickBird multispectral imagery through band Ratios and fuzzy comparison. *ISPRS Journal of Photogrammetry & Remote Sensing*, 64(1), pp. 55-64.

Yamamoto, T., Hanaizumi, H. and Chino, S., 2001. A change detection method for remotely sensed multispectral and multitemporal images using 3-D segmentation. *IEEE Transactions on Geosciences and Remote Sensing*, 39(5), pp. 976-985.

Yang, Y., Han, C. and Han, D., 2008. A Markov random field model-based fusion approach to segmentation of SAR and optical images. In: *Proccedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 4, pp. 802-805.

Zadeh, L. A., 1965. Fuzzy Sets. *Information and Control*, 8(3), pp. 338-353.

Zhang, Y. J., 1997. Evaluation and comparison of different segmentation algorithms. *Pattern Recognition Letters*, 18(10). pp. 963-974.

Zhong, C., Zhongmin, Z., DongMei, Y. and Renxi, C., 2005. Multi-scale segmentation of the high resolution remote sensing image. In: *Proccedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 5, pp. 3682- 3684.

ACKNOWLEDGEMENTS

This research is funded by the Strategic Project Grants Program of the Natural Science and Engineering Research Council of Canada (NSERC).

ASSESSMENT SYSTEM OF GIS-OBJECTS USING MULTI-TEMPORAL IMAGERY FOR NEAR-REALTIME DISASTER MANAGEMENT

D. Frey*, M. Butenuth

Remote Sensing Technology, Technische Universität München,
Arcisstr. 21, 80333 München, Germany – daniel.frey@bv.tum.de, matthias.butenuth@bv.tum.de

KEY WORDS: Change Detection, Classification, Integration, Multitemporal, Multisensor, Fusion, Disaster

ABSTRACT:

In this paper, a damage assessment system of GIS-objects such as roads and buildings after natural disasters is presented. The main contribution is the integration and exploitation of multi-temporal imagery leading to a more robust assessment of infrastructural objects. In addition, the chronological development of the assessed objects is investigated. The multivariate alteration detection method is used to detect changes between different time points in conjunction with the classification of different changes realized via Gaussian mixture models. Further accessorily introduced information are derived from GIS, in particular DEM belief functions. The strategy of the proposed approach is the combination of the computed probabilities using individual appropriate methods. The goal of the system is the assignment of GIS-objects into different damage assessment categories as intact or not intact/destroyed using the fused information from multi-temporal multi-sensorial data. The system is tested at a test scenario assessing roads concerning their trafficability. The results show the improvement of the damage assessment system after the integration of multi-temporal information.

1. INTRODUCTION

In this paper, an assessment system of GIS-objects is presented using multi-sensorial and multi-temporal imagery after natural disasters. The focus of this article is the multi-temporal component, because the integration of imagery from different time points into an assessment system has several advantages: Firstly, multi-temporal images provide the opportunity to monitor natural disaster chronologically during a period of time, not only at a specific time point. Secondly, the assessment of the GIS-objects at the time point t_2 can be improved using the results from time point t_1.

Another focus of this article is the automatic information retrieval from imagery being relevant for rescue teams after natural disasters. Information on the status of the infrastructure after disasters is essential to guarantee an effective and fast disaster management. Therefore, the emphasis of this article is the development of automated methods assessing infrastructural objects such as roads concerning their functionality.

The precondition ensuring an effective disaster management is the near-realtime supply of information, because time is the crucial parameter. Therefore, great efforts have been made in order to speed up the workflow from satellite tracking and data acquisition to the point of map generation (Voigt et al., 2007). The whole workflow can generally be passed within 24 hours. Data analysis consisting of information extraction, damage assessment, thematic analysis and change detection plays a decisive role in the processing chain of the workflow (Bamler et al., 2005). Up to now a lot of data analysis tasks are done manually which is very time consuming. Therefore, automation is required to substitute the manual interpretation. The difficulty is the development of methods minimizing wrong decisions to avoid fatal consequences in emergency actions. Possibilities to achieve a low error rate are semi-automated approaches.

A given fact is the variability of available imagery and GIS data in case of emergency. For this reason, a basic characteristic of the presented system is the handling of different input data sources. In this article, a modular system is presented which is able to deal with varying data sources and provides the embedding of all available information.

In Section 2. existing up-to-date damage assessment systems are presented and categorized in *area-* and *object-based systems*. In addition, data fusion techniques with regard to disaster management are discussed. Hereupon, the basics of Gaussian mixture model and a the change detection methods are introduced since these methods are key elements of the assessment system, which is described in Section 3. In Section 4. the general system is applied to a test scenario, the shown results are evaluated concerning their quality measure. Finally, further investigations and future work is pointed out.

2. STATE OF THE ART AND BASICS

2.1 Damage Assessment Systems

In case of natural disaster it is reasonable to differentiate between *object-based* and *area-based* damage assessment systems. The focus of *object-based systems* is the assessment of infrastructural objects such as roads or buildings concerning their functionality. In recent years several systems have been developed estimating the extent and type of destruction on various buildings. The damage assessment was realized using different kind of sensors such as LIDAR (Rehor et al., 2008) or satellite images (Chesnel et al., 2007). But there is a lack of methods assessing transportation lifelines after natural disasters (Morain and Kraft, 2003). In (Frey and Butenuth, 2009) a near-realtime assessment system of roads using GIS-objects and multi-sensorial data is presented. The road objects are classified into different states and are visualized using the ample paradigm proposed by Förstner (Förstner, 1996). In this article, the system is extended by the multi-temporal component using change detection methods.

On the other hand *area-based systems* focus on the affected regions. Typical examples are the generation of flood masks derived from different sensors. Besides optical imagery, particularly radar images are suitable for the extraction of inundated areas. Martinis (Martinis et al., 2009) uses a split-based automatic

* Corresponding author.

thresholding method to detect flooded areas from TerraSAR-X data in near real-time.

2.2 Data fusion

In general, the performance of the damage assessment system can be improved by adding additional imagery and data sources. The additional benefit depends on the way of how the data is combined. Pohl (Pohl and Van Genderen, 1998) differentiate between three different levels of image fusion: pixel level, feature level and decision level. The combination of different data sources, e.g. vector and image data, was discussed in several other contributions, e.g. (Butenuth et al., 2007). Particularly, the integration of GIS information combined with imagery improves the results and simplifies the decision makings enormously (Brivio et al., 2002). Wang (Wang et al., 2002) presents a method for mapping flood extend combining optical imagery and DEM. In the approach, for each data source an individual flood mask is generated. The final flood mask consists of the set union of the individual masks. Considering the DEM as an image this approach belongs to the decision level image fusion as defined in (Pohl and Van Genderen, 1998). The presented approach in this article combines imagery and DEM, too, to detect flooded areas. In contrast to the discussed approaches, the aim is the combination based on probabilities derived from the input data.

2.3 Change Detection: Multivariate Alteration Detection (MAD)

Change detection algorithms are widely used investigating the extent and damage of natural disasters. A comprehensive review about change detection methods is given in Lu (Lu et al., 2004). However, many methods are restricted to specific sensors characteristics. The efficient response in case of natural disaster requires a change detection method which is able to deal with various sensors containing a different number of channels. Furthermore, the influence of changing atmospheric conditions should be minimized. The multivariate alteration detection method (MAD) is invariant to linear transformations which implies the insensitivity to linear atmospheric conditions or sensor calibrations at two different times. In addition, the handling of different numbers of channels is given (Nielsen et al., 1998).

The MAD transformation is based on the canonical correlation analysis (CCA). The CCA investigates the intercorrelation between two sets of variables unlike the principal component analysis, which identifies patterns of relationship within one set of data. Let $\mathbf{F} = \{F_1, F_2, ..., F_n\}$ and $\mathbf{G} = \{G_1, G_2, ..., G_m\}$ being two images with n or m channels ($n \leq m$). A linear combination of the intensities for all channels leads to the transformed images *U* and *V*:

$$\begin{aligned} U &= \vec{a}\,\mathbf{F} &&= a_1F_1 + a_2F_2 + \ldots + a_nF_n \\ V &= \vec{b}\,\mathbf{G} &&= b_1G_1 + b_2G_2 + \ldots + b_mG_m. \end{aligned} \quad (1)$$

The goal of the transformation is to choose the linear coefficient $\vec{a}$ and $\vec{b}$ minimizing the correlation between *U* and *V*. This leads to the result that the difference image between the transformed images *U* and *V* will have maximum variance. Due to the fact that multiples of *U* and *V* would have the same correlation a reasonable constraint $var(U) = 1$ and $var(V) = 1$ is chosen:

$$\begin{aligned} var(U - V) &= var(U) + var(V) - 2cov(U, V) \\ &= 2(1 - cov(U, V)). \end{aligned} \quad (2)$$

Using CCA, the linear coefficients $\vec{a}$ and $\vec{b}$ are determined and the MAD variates M_i can be calculated (Nielsen et al., 1998):

$$M_i = U_i - V_i \quad \text{for } i = 1 \ldots n. \quad (3)$$

An extension to the MAD transformation is the iteratively reweighted MAD (IRMAD) method. Similar to boosting methods in data mining, an iteration schema focuses on observations whose change status is uncertain (Nielsen, 2007). Since the MAD or IRMAD variates can only being interpreted in a statistical manner there is a need to assign semantic meaning to the MAD variates. In Canty (Canty and Nielsen, 2006) an unsupervised classification method is proposed based on the MAD variates to cluster pixel in no-change and one or more change categories.

2.4 Combination of Probability Functions: Gaussian Mixture Model (GMM)

Since the radiometric characteristics of infrastructural objects of the same type could vary strongly, single probability functions are not able to describe the complex scenes sufficiently. Therefore, mixture models which combines single functions to a more complex probability function are used. The resulting probability function $p(\mathbf{y}|\theta_j)$ is simply a weighted sum of the initial functions $p(\mathbf{y}|\theta_j)$:

$$p(\mathbf{y}|\theta) = \sum_{j=1}^{k} \alpha_j p(\mathbf{y}|\theta_j). \quad (4)$$

Each θ_j describes the set of parameters defining the jth component, $\alpha_1 \ldots \alpha_j$ are the weights called mixing probabilities and $\mathbf{y} = [y_1 \ldots y_d]^{\mathrm{T}}$ represent one particular outcome of a d-dimensional random variable $\mathbf{Y} = [Y_1 \ldots Y_d]^{\mathrm{T}}$. If $\mathbf{Y}$ is normal distributed, Gaussian are typically used. The mixing probabilities have to fulfill following equations:

$$\alpha_j \geq 0, \quad j = 1 \ldots k, \quad \text{and} \quad \sum_{j=1}^{k} \alpha_j = 1. \quad (5)$$

The expectation maximization (EM) algorithm is used to determine α_j and θ_j. A detailed description of mixture models can be found in McLachlan (McLachlan and Peel, 2000). The minimum message length criterion (MML) is one possibility to find the number of centers j and is used in our system (Figueiredo and Jain, 2002).

3. ASSESSMENT SYSTEM

In this Section, the general assessment system is presented using multi-sensorial multi-temporal imagery and further available data. The goal is the assessment of GIS-objects categorizing them into different states.

3.1 System

The design of the system has a modular and very flexible structure to cope with varying raw data being available in emergency cases (cf. Fig. 1). Nevertheless, there are some prerequisites to apply the system. The GIS-objects which should be assessed concerning their functionality must be given. It is conceivable to extract the GIS-objects using imagery before the natural disaster takes place or, alternatively, from a GIS. However, in view of the performance of automatic extraction methods, objects from a given GIS-database with a guaranteed quality are better suited. The result of the assessed GIS-objects depends strongly on the

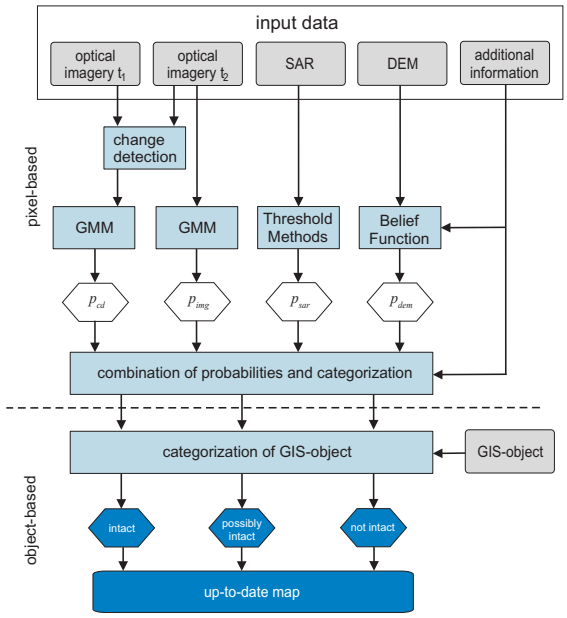


Figure 1: General damage assessment system.

available input information. Besides the imagery, DEM and further GIS-information can be embedded into the system. Here, this data is called input data.

For multispectral imagery Gaussian mixture models are applied. Belief functions are introduced to derive probabilities from GIS-information. If multi-temporal imagery are available change detection methods such as the MAD algorithm are used to derive probabilities. The combination of the different input data is carried out in the probability level. All the individual methods and the combination of the probabilities are realized at pixel level. In contrast, the subsequent assignment of GIS-objects to the categories *intact*, *possibly intact* or not *intact/destroyed* using a maximum likelihood estimation is object-based (cf. Fig. 1).

3.2 Methods and Combination of Probabilities

For each input data individual methods have to be applied to derive individual probabilities if the infrastructural objects are intact or not (cf. Fig. 1). Given multispectral imagery as input data a multispectral classification is carried out. The infrastructural objects are classified to different classes relating to the categories *intact*, *possibly intact* and not *intact/destroyed*. Since a lot of classes like roads have no consistent radiometric characteristic as shown in Figure 2 and Figure 3, GMM are used to deal with the different subgroups of the classes. The resulting probabilities from the mixture model p_{img} are combined with probabilities from further input data (cf. Fig. 1).

The availability of images at different time points enables the usage of change detection methods exploiting additional assessment criteria. The IRMAD algorithm enables the detection of changes caused by natural disasters. The resulting IRMAD-variates are classified using a supervised multispectral classification. For the different change-classes, i.e. 'intact $\Rightarrow$ destroyed' probability functions are generated. These probabilities p_{mad} are embedded into the assessment system. In Figure 4(c) three IRMAD variates are shown as an RGB-color image obtained from IKONOS-images at time t_1 (cf. Fig. 4(a)) and time t_2 (cf. Fig. 4(b)). In this example of a flood event the changed areas from flooded to not flooded are illustrated in pink, the gray color stands for no change (cf. Fig. 4(c)).

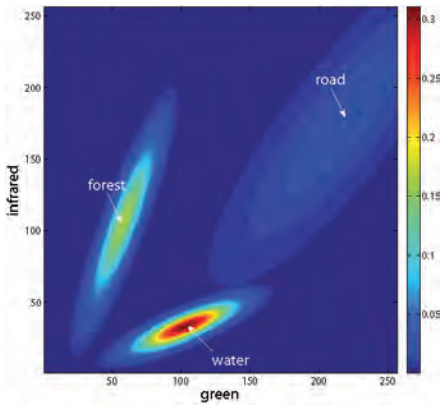


Figure 2: Two-dimensional probability density functions of the classes forest, water and a combined class road. Exemplarily visualized via the infrared and green channel.

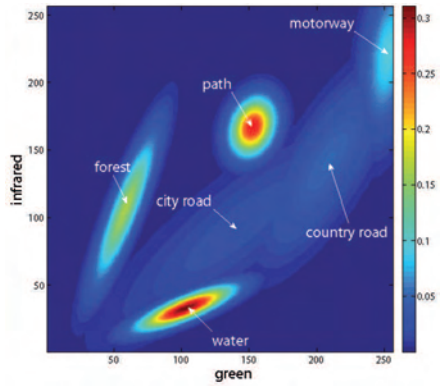


Figure 3: Two-dimensional probability density functions of the classes forest, water and separated road-classes (city road, country road, path and motorway). Exemplarily visualized via the infrared and green channel.

Additional GIS-information such as DEM is often available having the opportunity to enhance the assessment system. Since the combination of the input data is based on the probability level, also from the GIS-information probabilities have to be derived. Belief functions can be generated depending on the GIS-information. In Figure 5 an example is shown, which depicts the probability that an object is flooded depending on the altitude. The combination of the probabilities derived from the different input data is defined as following (cf. Fig. 1):

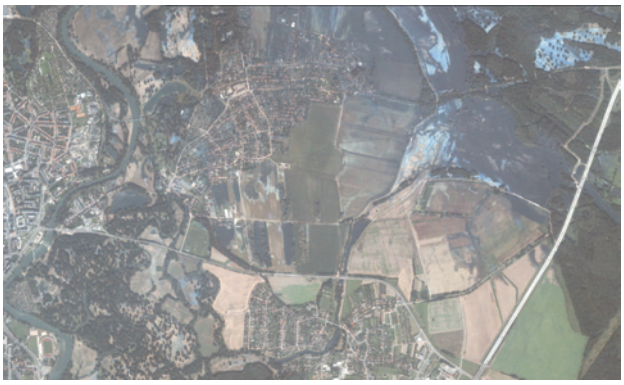
$$\begin{aligned} p_{s_1} &= p_{s_1,img} \otimes p_{s_1,gis} \otimes \ldots \otimes p_{s_1,mad} \\ p_{s_2} &= p_{s_2,img} \otimes p_{s_2,gis} \otimes \ldots \otimes p_{s_2,mad} \\ &\vdots \\ p_{s_i} &= p_{s_i,img} \otimes p_{s_i,gis} \otimes \ldots \otimes p_{s_i,mad}. \end{aligned} \tag{6}$$

The probabilities p_{s_i} are the combined probabilities of one status s_i. In the easiest case the set of states could be *intact* or *not intact*. But it is also possible to think of different kinds of destruction states. In addition, weights are introduced since the information content of the different input data varies:

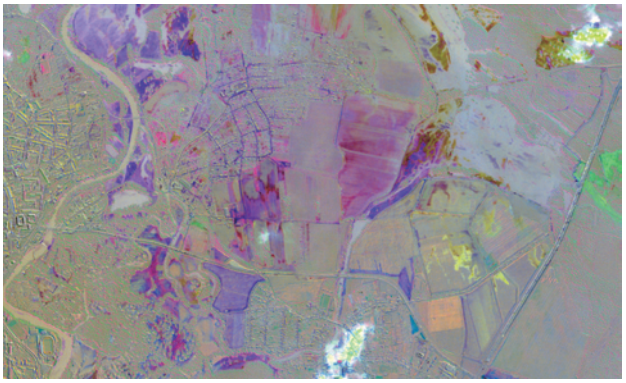
$$p_{s_i} = w_1 p_{s_i,img} \otimes w_2 p_{s_i,gis} \otimes \ldots \otimes w_d p_{s_i,mad}. \tag{7}$$

The number of input data is denoted as d. Finally, the object is categorized to the state s_i with the largest probability.

(a) IKONOS-scene of flooded area at time t_1



(b) IKONOS-scene of flooded area at time t_2



(c) Three MAD-variates depicted as an RGB-color image

Figure 4: Change detection using MAD-algorithm.

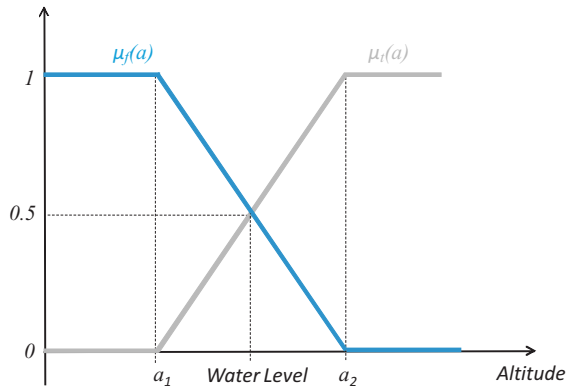


Figure 5: Belief functions depending on altitude: area is flooded (blue), area is not flooded (gray).

4. RESULTS AND DISCUSSION

The presented damage assessment system is applied to a specific flood scenario. In real case scenarios the availability of input data is the crucial factor. The derivation of the probabilities given in Equation 6 is not always possible depending on the available data. On the other side often additional information exist which are useful to generate additional rules. In real applications the combination of probabilities is embedded into a rule-based framework which can differ from case to case.

4.1 Test Scenario

Test scenario is the flooding of the river Elbe (Germany) in the year 2002. The available input data for the damage assessment system consists of two IKONOS-scenes (cf. Fig. 4(a) and 4(b)) acquired at the 21th and 26th of August, and a DEM. The peak of the water level was measured at the 19th of August. The scene at the time t_1 shows almost the maximum inundated area. In the second scene at time t_2 the flooding receded strongly and only a small area is covered by water (cf. Fig. 4(b), top right). In addition to the images, a DEM is available with a 10m x 10m grid with an geometric accuracy of +/- 1m. In this test scenario road objects given from a GIS-database are assessed concerning their trafficability.

4.2 Workflow of Rule-based Classification

A detailed workflow of the rule-based assessment system is depicted in Figure 6, the input data are illustrated by gray parallelograms. Below these parallelograms the derived probabilities from the input data are attached in gray rectangles. The combination of the probabilities is realized in the blue boxes. The goal in this scenario is the assessment of road segments concerning the trafficability at time t_2. In addition to the imagery and the DEM described in Section 4.1, the assessed road segments at the time t_1 are given. They are obtained by means of the described assessment system using very strict parameters. Alternatively, also a manual generated reference at time t_1 could be used. The assessed road segments at time t_1 and additional information as the water level lead to the rule-based framework built on the combination of the probabilities. The probability p_{img} derived from the imagery is partitioned into three different probabilities belonging to a specific class: water p_{water}, road p_{road}, forest p_{forest}. As shown in Figure 3 the class road is subdivided into subclasses of roads using GMM. Using a maximum likelihood estimation followed by a threshold operation the segment is categorized into the three states *trafficable*, *possibly flooded* and *flooded*.

4.3 Evaluation

The obtained results are compared to a manually generated reference. The information for the generation of the reference is only the image at time t_2. Therefore, it is not a comparison with the real ground truth, but it is the comparison of the automatic approach with the manually interpretation of an human operator. The reference is also categorized into three different classes *trafficable*, *possibly flooded* and *flooded*. Since the categorization of the automatic system consists of the same states the following four different assignment criteria are determined: 'correct assignment', 'manual control necessary', 'possibly correct assignment' and 'wrong assignment'. The category 'correct assignment' means that the manually generated reference is identical with the result of the automatic system. In the case of 'manual control necessary' the automatic approach leads to the state *possibly flooded* whereas the manual classification assigns the line segments to *flooded* or *trafficable*. The other way around

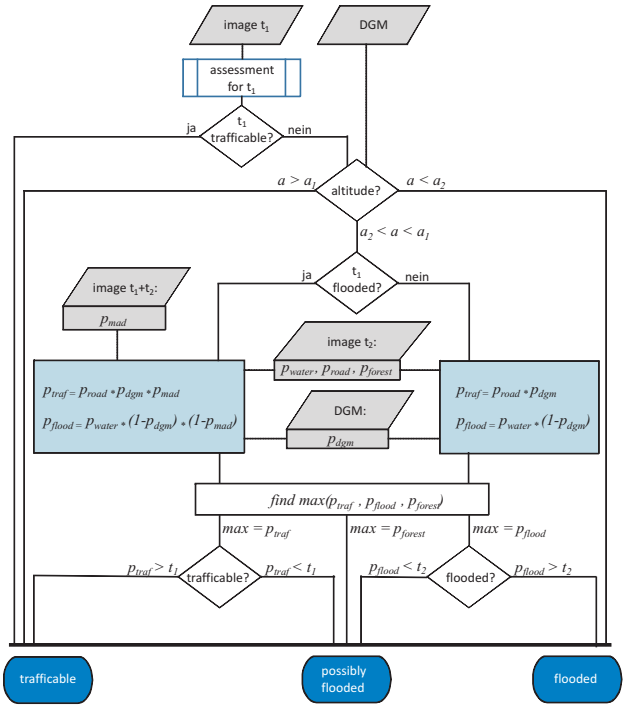


Figure 6: Workflow of Rule-based Classification.

denotes the expression 'possibly correct assignment'. 'wrong assignment' means that one result categorize the segment to *flooded* and the other to *trafficable*. The enhancement of the automatic system by the combined interpretation is shown in Table 7.

	t_2	t_2, DEM	$t_{1,2}, DEM$	$t_{1,2,c}, DEM$
correct	68.40	68.45	69.60	87.14
manual	27.88	27.77	27.48	10.96
possibly	2.64	2.72	2.52	1.79
wrong	1.08	1.06	0.40	0.11

Table 7: Evaluation (percentage shares)

The first column in Table 7 represents the result using only the image t_2 without any further information. The result with about 1% of 'wrong assignements' and about 68% 'correct assignment' is almost identical if an additional DEM as input data is used (cf. Table 7: t_2, DEM). The reason for the lack of improvement could be ascribed to the bad accuracy of the used DEM. The influence of the height information is discussed in (Frey and Butenuth, 2009). The evaluated road segments are depicted in Figure 8(a). Green road segments correspond to 'correct assignment', yellow to 'manual control necessary', cyan to 'possibly correct assignment' and red or blue belongs to 'wrong assignment'. If the systems assigns a road segment to the category *trafficable* but the referencce is *flooded* the road segment is colored in red. Blue road segments are assigned to *flooded* by the system and *trafficable* by the reference.

In Figure 8(b) the result of the third column from Table 7 is visualized which includes the additional scene at time point t_1 as input data. The additional scene and the resultant calculated probability p_{mad} derived from the described MAD method leads to an improvements of the results. Several red road segments disappear whereas the 'correct assignments', the assignments to 'manual control necessary' and the 'possibly correct assignments' remains almost constant.

In Figure 8(c) the results exploiting an additional manually generated reference from scene t_1 are plotted. The numerical evaluation is presented in the forth column of Table 7 ($t_{1,2,c,DEM}$). The results are by far better then the previous obtained results. The 'correct assignments' arise from 69% to 87% and the 'wrong assignments' decrease from 0.4% to 0.1%. But it is important to point out, that a correct reference at the time point t_1 has to be generated. Nevertheless, it has no influence of the fact that the system is near-realtime since the time consuming generation of the reference can be done before.

(a) Detail of evaluation using image t_2 and DEM

(b) Detail of evaluation using image t_2, image t_1 and DEM

(c) Detail of evaluation using image t_2, image t_1 with correctly assessed roads and DEM

Figure 8: Evaluation of assessment system: green='correct assignment', yellow='manual control necessary', cyan='possibly correct assignment', red='wrong assignment'[system = *trafficable*, reference = *flooded*], dark blue = 'wrong assignment' [system = *flooded*, reference = *trafficable*].

4.4 Result after Data Fusion

The final obtained result using the described damage assessment system is depicted in Figure 9. All road segments are divided

into four different categories. Besides the already mentioned categories *trafficable* (green), *possibly flooded* (yellow) and *flooded* (red) an additional category *flooded to trafficable* (blue) is introduced by means of the change detection algorithm. This additional category is very useful for rescue teams since it shows the areas which are again trafficable after flooding.

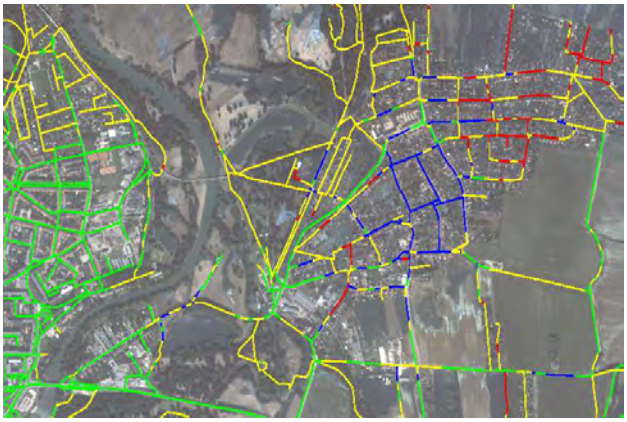


Figure 9: Detail of result of damage assessment system using all available input data: image t_1, imaget_2, DEM and manual generated reference at time t_1. (green = *trafficable*, yellow = *possibly flooded*, red = *flooded*, dark blue = *flooded* $\Rightarrow$ *trafficable*).

5. CONCLUSIONS

In this article, the general framework of a damage assessment system and the benefit of the included data fusion is shown. The improvement of the results by adding additional available data is demonstrated in the test scenario. The integration of multitemporal imagery leads to an enhancement of the damage assessment system concerning the correctness of the assessed objects and concerning the additional temporal information which can provide the rescue teams in emergency actions. Combining this basis with rule-based approaches which are strongly dependent on the natural disasters and available input data the overall system leads to useful results with a very little rate of 'wrong assignments'.

In future work, the generic system will be tested at more test scenarios with different sensors. In particular, the combination of optical images and radar images should be investigated in more detail. In addition, the influence of the DEM accuracy has to be investigated in future work. Besides the radiometric exploitation of the optical imagery also the geometric features should be introduced as an additional evidence of destructions. A distinction between different regions of global context should improve the results as well. Depending on the global context the required parameter can be chosen. The automatic setting of the parameters in the system is currently not included in this paper. Further investigations have to be done to learn suitable parameters automatically depending on the available data.

REFERENCES

Bamler, R., Reinartz, P., Riedlinger, T. and Schroeder, M., 2005. Moderne Raumfahrttechniken für Prävention bei Naturkatastrophen und das Krisenmanagement.

Brivio, P., Colombo, R., Maggi, M. and Tomasoni, R., 2002. Integration of remote sensing data and GIS for accurate mapping of flooded areas. International Journal of Remote Sensing 23(3), pp. 429–441.

Butenuth, M., Gösseln, G., Tiedge, M., Heipke, C., Lipeck, U. and Sester, M., 2007. Integration of heterogeneous geospatial data in a federated database. ISPRS Journal of Photogrammetry and Remote Sensing 62(5), pp. 328–346.

Canty, M. and Nielsen, A., 2006. Visualization and unsupervised classification of changes in multispectral satellite imagery. International Journal of Remote Sensing 27(18), pp. 3961–3975.

Chesnel, A.-L., Binet, R. and Wald, L., 2007. Quantitative assessment of building damage in urban area using very high resolution images. In: Urban Remote Sensing Joint Event, 2007, pp. 1–5.

Figueiredo, M. and Jain, A., 2002. Unsupervised learning of finite mixture models. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(3), pp. 381–396.

Frey, D. and Butenuth, M., 2009. Classification system of GIS-objects using multi-sensorial imagery for near-realtime disaster management. In: International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVIII(3/W4), pp. 103–108.

Förstner, W., 1996. Pros and cons against performance characterization of vision algorithms. In: Proceedings of ECCV Workshop on Performance Characteristics of Vision Algorithms, Cambridge, UK, April, pp. 215–218.

Lu, D., Mausel, P., Brondizio, E. and Moran, E., 2004. Change detection techniques. International Journal of Remote Sensing 25(12), pp. 2365–2401.

Martinis, S., Twele, A. and Voigt, S., 2009. Towards operational near real-time flood detection using a split-based automatic thresholding procedure on high resolution TerraSAR-X data. Natural Hazards and Earth System Science 9(2), pp. 303–314.

McLachlan, G. and Peel, D., 2000. Finite mixture models. Wiley-Interscience.

Morain, S. and Kraft, W., 2003. Transportation lifelines and hazards: Overview of remote sensing products and results. In: Proceedings of Remote Sensing for Transportation 29, pp. 39 – 46.

Nielsen, A., 2007. The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data. IEEE Transactions on Image Processing 16(2), pp. 463–478.

Nielsen, A., Conradsen, K. and Simpson, J., 1998. Multivariate alteration detection (MAD) and MAF postprocessing in multispectral, bitemporal image data: New approaches to change detection studies. Remote Sensing of Environment 64(1), pp. 1–19.

Pohl, C. and Van Genderen, J., 1998. Multisensor image fusion in remote sensing: concepts, methods and applications. International Journal of Remote Sensing 19, pp. 823–854.

Rehor, M., Bähr, H., Tarsha-Kurdi, F., Landes, T. and Grussenmeyer, P., 2008. Contribution of two plane detection algorithms to recognition of intact and damaged buildings in lidar data. The Photogrammetric Record 23(124), pp. 441–456.

Voigt, S., Kemper, T., Riedlinger, T., Kiefl, R., Scholte, K. and Mehl, H., 2007. Satellite image analysis for disaster and crisis-management support. IEEE Transactions on Geoscience and Remote Sensing 45(6), pp. 1520–1528.

Wang, Y., Colby, J. and Mulcahy, K., 2002. An efficient method for mapping flood extent in a coastal floodplain using landsat tm and dem data. International Journal of Remote Sensing 23(18), pp. 3681–3696.

THE POTENTIAL OF A SPACE ALTIMETER WHEN MEASURING SIGNIFICANT WAVE HEIGHT

B.Y. Grishechkin [a, *], A.I. Baskakov [a]

Dept. of Radiotechnical Devices, Moscow Power Engineering Institute (Technical University), Russia, 111250, Moscow, Krasnokazarmennaya street, 14
bg1722@aol.com, baskakovai@mpei.ru

KEY WORDS: Radar, Correlation, Measurement, Method, Sea

ABSTRACT:

This paper is dedicated to a new method of sea surface significant wave height (SWH) measuring. The method can be applied to space-born radar altimeter data. This method is based on the fundamentals of Bayesian optimal measuring of a parameter. The correlation function of sea surface echo was used to calculate a state of the art likelihood function of the SWH estimate. Optimal discriminator of SWH is developed using the least square error of the SWH estimate criteria, apart from others papers which mostly minimize the least square error of altimeter waveform fitting function. Discriminator characteristics are plotted and subsequent results are presented as well as potential measurement accuracy calculation results.

1. INTRODUCTION

Circular orbit space-born radar altimeter (RA) is used to produce broad field information, e.g. for monitoring of mean sea level and sea currents, refining of sea geoid, mapping of sea rings and gravitational anomalies, assessing of sea wave height and wind, monitoring of sea ices and rains, refining of Global Ocean model, some natural gas and mineral oil seeking activities at the sea shore, planning of sea freight and fishing, etc [Bhatt, 2005; Davis, 1997; Graham, 1998].

Optical and radar tools are used to measure an altimeter vehicle position with centimetre level error [Blanc et al, 1995]. RA emits C and Ku bands pulses to sea surface vertically, usually to illuminate a circle several kilometres in diameter [Chelton et al, 1989], Ka band will probably be used in the future.

Nowadays RA illuminate radio pulses with duration τ_p from tens to hundred microseconds and pulse power up to tens watts. Also linear frequency modulation is used with up to Δf_s=320 MHz spectrum bandwidth (480 MHz will probably be used in the future), i.e. so-called chirp pulse. The spectrum bandwidth is restricted not only by technical issues, but also by the effect of ionosphere dispersion [Min Ho Ka and Baskakov, 2004].

Parabolic reflector receive-transmit antenna is usually used for RA with a beam width of 1°-2°, and the maximum permissible antenna pointing error is one quarter of the beam width.

Apparently world-famous paper dedicated to the first space altimeter experiment was [McGoogan et al., 1974] and then probably the most referenced paper discussing the echo model was [Brown, 1977] followed by [Hayne, 1980].

There are known algorithms of retrieving geophysical parameters, including SWH, from RA data. Some of them involve intuitive evident method of assessing of altimeter waveform leading edge slope [Chelton et al, 1989; Tucker, 1991; Sujit Basu and Pandey, 1991], others involve maximum likelihood estimation [Rodriguez and Chapman, 1989; Challenor et al., 1990] which can actually be called a fitting algorithm. But there is not enough information about using the fundamental Bayesian optimal method though a very brief paper was published [Mailhes et al, 2008].

This paper describes a developed algorithm of SWH estimation

2. CORRELATION FUNCTION OF ECHO

By applying the Gaussian facet model of sea surface one can see that the echo is subjected to fast and slow fluctuations. The fast fluctuations of echo signal occur when RA-emitted radio wave leading edge crosses sea surface facets, which provides random initial phases of echoes. The slow fluctuations occure when the reflecting area increases during the sensing pulse. It was shown [Baskakov et al, 2007] that correlation interval of the echo fast fluctuations at the input of the RA receiver is inversely proportional to the compressed pulse duration (almost during the whole duration) therefore the de-chirp technique of receiving may be used as an optimal in the signal to noise ratio criteria meaning. The echo slow fluctuations calculation [Min Ho Ka and Baskakov, 2004] allowed the selection of sensing pulse repetition interval. It was shown [Baskakov et al, 2007] that for space-born altimeter correlation function of RA echo at the matched filter (MF) output can be presented as a product of two independent functions

$$R(t,\tau) = \mathrm{Re}\left[P_m \rho(\tau) P_n(t) \exp(-j\omega_0 \tau)\right] \qquad (1)$$

where P_m – maximum power of the echo at the output of the MF, P_n – normalized by P_m averaged form of the power of the signal at the MF output, $\rho(\tau)$ – a fast changing function in comparison with $P_n(t)$ differing from the chirp auto-correlation function by an exponent, ω_0 – centre frequency of RA sensing pulse.

Theoretically the RA receiver can be based on a MF, processing the echo in the time domain, or based on a de-chirp scheme, processing the echo in the frequency domain. These two schemes are equal in theory, but in practice the second variant is deployed more. In the time domain, the MF output signal power can be presented [Baskakov et al, 2007] as

$$P_n(t) = \exp\left[\frac{b^2\mu^2}{4} - bt\right] \cdot \left[0{,}5 - \Phi_1\left(\frac{b\mu}{2} - \frac{t}{\mu}\right)\right] \qquad (2)$$

* Corresponding author

where b – parameter depending on the antenna beam width, antenna pointing error and sea surface roughness; μ – parameter depending to the sea surface wave height supposing that facets ordinates probability density is Gaussian and

$$\mu = \sqrt{\frac{1}{\pi \Delta f_s^{\,2}} + 2\left(\frac{2\sigma_z}{c}\right)^2} \qquad (3)$$

c – light velocity; σ_z – sea surface facets standard deviation, which can be recalculated into the significant wave height (SWH) $H_{1/3} = 4\sigma_z$ [Chelton, 1989; Baskakov et al, 2007]; $\Phi_1(x)$– probability integral divided by 2.

When processing the echo in the frequency domain the average power spectrum of the signal at the de-chirp mixer output accounting for side lobe correction can be calculated as Fourier transform of the two-dimension correlation function (1) and after normalizing by maximum value of power spectral density S_m can be presented as [10]

$$S_n(f) = \exp\left[\frac{b^2\mu^2}{4} - \frac{bf}{k}\right] \cdot \left[0{,}5 - \Phi_1\left(\frac{b\mu}{2} - \frac{f}{k\mu}\right)\right] \qquad (4)$$

where k – frequency slope in the sensing pulse and the frequency axis f is shifted by the intermediate frequency value f_{int}.

Altimeter waveform is the average power form of the echo at the MF output as well as the average power spectrum density at the de-chirp mixer output. The vehicle height above the sea surface, SWH, sea surface backscattering coefficient σ_0 and antenna pointing error relate to the offset and the shape of the waveform [Baskakov et al, 2007], therefore they can be revealed by processing the waveform. Practically the most demanding part of the process is fulfilled by on-Earth centres [Zigna et al, 2001].

3. OPTIMAL DETECTION

According to the theory of statistical radio location there must be a detection process before measuring. After resolving an according Fredholm equation the logarithm of likelihood function for RA was presented in [Grishechkin and Baskakov, 2007] as

$$\Lambda[\xi(t)/\tau_0] = C_0 - \frac{1}{2N_0}\int_0^T \xi^2(t)dt + \frac{1}{2N_0}\int_0^T \frac{q(t)}{1+q(t)}\left|\int_{-\infty}^{\infty} \xi(\tau) e^{j\omega_0\tau}\, \dot{h}_f(t-\tau)d\tau\right|^2 dt \qquad (5)$$

where ξ(t) – input signal, $\dot{h}_f(t-\tau)$ – pulse response characteristic of the MF, q(t) – current ratio of the echo power to the power spectral density of the RA receiver noise (N_0) at the MF output, T – duration of the echo, C_0 – a constant which does not depend on the echo.

4. OPTIMAL ESTIMATION

A discriminator is the key unit of any tracking measuring scheme. The main task for RA is to measure its altitude. Height discriminator analysis was presented in [Baskakov, 1994] and other papers. This paper presents analysis of SWH discriminator.

The output signal of an optimal (meaning the least square error of the estimated parameter) discriminator is determined by the measuring parameter derivative of the likelihood function logarithm (5). The optimal discriminator algorithm was mathematically calculated assuming that other parameters of the waveform were fixed. Then the influence of other parameters variation was assessed. Mathematically it was preferable to differentiate (5) by parameter μ (3) instead of SWH value. Parameter μ can be recalculated into standard deviation of sea surface facets ordinates σ_z and then into SWH as mentioned above.

As well as to processing waveform in the time domain [Grishechkin et al, 2008] it can be shown that for processing in the frequency domain the discriminator characteristic has the following formula

$$\overline{U_{d\mu}(\delta\mu)} = S_m \int_{-\infty}^{\infty} S_n(f,\mu) v_{r\mu}(f,\mu+\delta\mu)\mathrm{d}f \qquad (6)$$

where the discriminator reference function has the following formula

$$v_{r\mu}(f,\mu+\delta\mu) = \frac{\frac{1}{2}b^2(\mu+\delta\mu)S_n(f,\mu+\delta\mu)}{q_m\left[\frac{1}{q_m} + S_n(f,\mu+\delta\mu)\right]^2} - \frac{\exp\left[\frac{b^2(\mu+\delta\mu)^2}{4} - \frac{bf}{k}\right]\cdot\left[\frac{b}{2} + \frac{f}{k(\mu+\delta\mu)^2}\right]}{\sqrt{\pi}\, q_m\left[\frac{1}{q_m} + S_n(f,\mu+\delta\mu)\right]^2} \cdot \frac{\exp\left[-\left(\frac{b(\mu+\delta\mu)}{2} - \frac{f}{k(\mu+\delta\mu)}\right)^2\right]}{\sqrt{\pi}\, q_m\left[\frac{1}{q_m} + S_n(f,\mu+\delta\mu)\right]^2} \qquad (7)$$

and the signal to noise ratio in the bandwidth of one filter of the spectrum processing unit corresponding to the highest point of RA waveform has the following formula

$$q_m = \frac{S_m}{N_0 \Delta F_c T} \qquad (8)$$

where ΔF_c is inversely proportional to the correlation interval of the fast fluctuations mentioned above.

Equations (6) and (7) were used to plot the series of discriminator characteristics, see figure 1. The characteristics are plotted as a function of δf error determined by the current altitude error. The 1 kHz step of δf corresponds to 10 cm step of

altitude. Other parameters are fixed: vehicle orbit height h = 1000 km, Δf_s=320 MHz, τ_p=102.4 μs, q_m = 20 db, true standard deviation sea surface facets ordinates $\sigma_{z\,tr}$ = 1 m.
Having (6) and (7) one can calculate the correlation function of the discriminator output signal and then calculate the spectrum of the output signal, which, at zero frequency, is equal to the standard deviation of the estimate of μ. The resulting standard deviation of the estimate of sea surface facets ordinates σ_z, which can be recalculated to SWH, is plotted in figure 2.

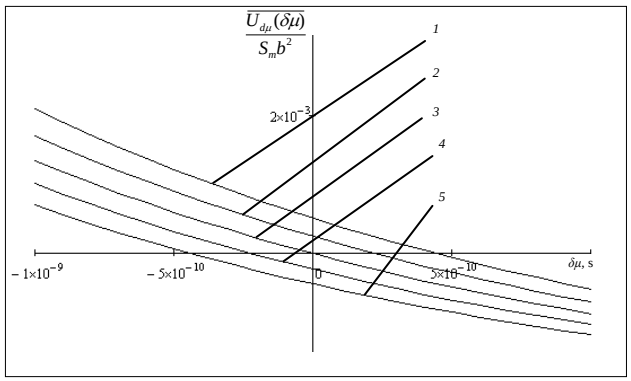


Figure 1. Sea wave height estimator discriminator characteristic series, 1 – δf=-2 kHz; 2 – δf =-1 kHz; 3 – δf =0 kHz; 4 – δf =1 kHz; 5 – δf =2 kHz

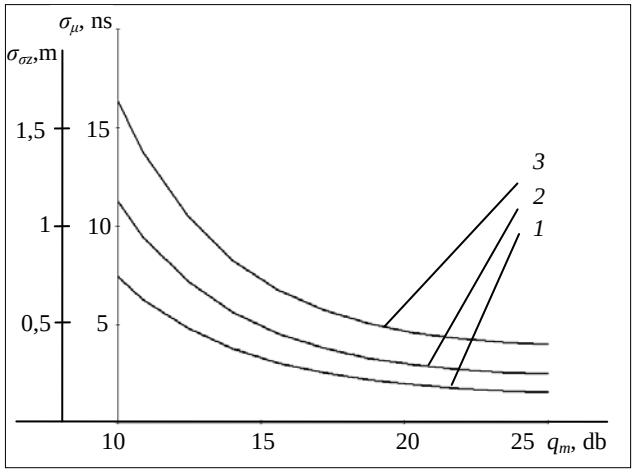


Figure 2. Potential accuracy of single measurement of parameters μ and σ_z,
1 – $\sigma_{z\,tr}$ = 1 m; 2 – $\sigma_{z\,tr}$ = 2 m; 3 – $\sigma_{z\,tr}$ = 4 m

Usually homogeneous sea surface roughness area is about 100 km or more, therefore it is adequate to average SWH measurements for about 4 seconds. If the pulse repetition frequency is 1 kHz and $\sigma_{z\,tr}$ is from 1 m to 4 m and signal to noise is 20 db then averaging for 4 s can provide less than 1 cm error of σ_z measurement, i.e. 4 cm error for SWH.

5. CONCLUSION

This paper and our previous papers reach the following conclusions:
1. Systematic error of optimal RA discriminators for SWH measurements depends on altimeter altitude error.
2. Optimal discriminator demands compensation of noise component of the discriminator output signal which can be provided by processing of RA input "signal" when there is no echo (e.g. before emitting the next pulse).
3. SWH measurements accuracy can not be reduced lower then the limit determined by the natural irreducible fluctuations of RA echo which caused by the random character of sea surface reflection. As expected, optimal discriminator for SWH estimator developed by using the Bayesian method is the most accurate of all the published.

REFERENCES

Baskakov A.I., 1994. *Precise Space-born Oceanographic Radar Altimeter: Students Manual*, Power Engineering Institute (Technical University), Moscow, Russia (in Russian).

Baskakov, A.I., Vazhenin, N.A., Morozov, K.N., 2000. Comparison of information signal processed using time- and frequency-domain methods in oceanographic precision radar altimeters. *Earth Observation Remote Sensing*, 16, pp. 449-455.

Baskakov, A.I., Min Ho Ka, Vazhenin, N.A, Grishechkin, B.Y., 2007. Correlation function analysis of precise space altimeter echo. *Proceedings of Universities in Russia. Radio electronics*, 1, pp. 53-62 (in Russian).

Bhatt, V., Kumar, R., Basu, S., Agarwal, V.K., 2005. Assimilation of altimeter significant wave height into a third-generation global spectral wave model. *IEEE Transactions on Geoscience and Remote Sensing*, 43(1), pp. 110-117.

Blanc, F., Le Traon, P.Y., Houry, S., 1995. Reducing orbit error for a better estimate of oceanic variability from satellite altimetry. *Journal of Atmospheric and Oceanic Technology*, 12(2), pp.150-160.

Brown, G.S., 1977. The average impulse response of a rough surface and its applications, *IEEE Transactions on Antennas and Propagation*, AP-25 (1), pp. 67-74.

Challenor, P.G., Srokosz, M.A., Tokmakian, R.T., 1990. Maximum likelihood estimation for radar altimetry, *Monitoring the Sea, IEE Colloquium on 18/01/1991*, pp. 10/1-10/3.

Chelton, D.B., Walsh, E.J., MacArthur, J.L., 1989. Pulse compression and sea level tracking in satellite altimetry. *Journal of Atmospheric and Oceanic Technology*, 6, pp. 407-438.

Davis, C.H., 1997. A robust threshold retracking algorithm for measuring ice-sheet surface elevation change from satellite radar altimeters. Geosci*ence and Remote Sensing, IEEE Transactions*, 35(4), pp. 974 – 979.

Graham, D., 1998. Determination of oceanic rain rate and rain cell structure from altimeter waveform data. Part I: Theory. *Journal of atmospheric and oceanic technology*, 15, pp. 1361-1378.

Grishechkin, B.Y., Baskakov, A.I., 2007. Echo detection characteristics of precise space born oceanographic radar altimeter. *Aerospace Devices Engineering*, 5, pp. 37-40 (in Russian).

Grishechkin, B.Y., Terehov, V.A., Baskakov, A.I., 2008. Analysis of systematic errors of optimal discriminator for sea wave heights measurements from space. *Radiotechnical copybooks*, 37, pp. 62-65 (in Russian).

Hayne, G.S., 1980. Radar altimeter mean return waveform from near-normal-incidence ocean surface scattering. *IEEE Transaction on Antennas and Propagation*, AP-28 (5), pp.687-692.

Mailhes, C., Tourneret, J.-Y., Severini, J, Thibaut, P., 2008. Bayesian estimation of altimeter echo parameters, *IEEE Internation Geoscience and Remote Sensing Symposium (IGARSS)*, Boston, USA.

McGoogan, J.T. Miller, L.S. Brown, G.S. Hayne, G.S., 1974. The S-193 radar altimeter experiment. *Proceedings of the IEEE*, 62(6), pp. 793-803.

Min Ho Ka, Baskakov, A.I., 2004. Restriction on the bandwidth of radio signals of a space-born precision radar altimeter due to the influence of the ionosphere. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E87-A(6), pp.1318-1322.

Rodriguez, E., Chapman, B., 1989. Extracting ocean surface information from altimeter returns: the deconvolution method. *Journal of Geophysical Research*, 94 (C7), pp. 9761-9778.

Sujit Basu, Pandey, P. C., 1991. Numerical experiment with modelled return echo of a satellite altimeter from a rough ocean surface and a simple iterative algorithm for the estimation of significant wave height, *Proc. Indian Acad. Sci. (Earth Planet. Sci)*, 100 (2), pp.155-163.

Tucker, M.J., 1991. Waves in ocean engineering, *Measurement, Analysis, Interpretation*. Ellis Horwood, Chichester, UK.

Zigna, JM., 2009. Poseidon 2 level 1.0 processing, http://podaac.jpl.nasa.gov/jason/documents/POS2_level10.pdf (accessed 22 Feb. 2010).

CLASSIFICATION OF SETTLEMENT AREAS IN REMOTE SENSING IMAGERY USING CONDITIONAL RANDOM FIELDS

T. Hoberg [a, *], F. Rottensteiner [a]

[a] IPI, Institute of Photogrammetry and GeoInformation, Leibniz Universitaet Hannover, Germany
(hoberg, rottensteiner)@ipi.uni-hannover.de

KEY WORDS: Conditional Random Fields, contextual information, classification, satellite imagery, urban area

ABSTRACT:

Land cover classification plays a key role for various geo-based applications. Numerous approaches for the classification of settlements in remote sensing imagery have been developed. Most of them assume the features of neighbouring image sites to be conditionally independent. Using spatial context information may enhance classification accuracy, because dependencies of neighbouring areas are taken into account. Conditional Random Fields (CRF) have become popular in the field of pattern recognition for incorporating contextual information because of their ability to model dependencies not only between the class labels of neighbouring image sites, but also between the labels and the image features. In this work we investigate the potential of CRF for the classification of settlements in high resolution satellite imagery. To highlight the power of CRF, tests were carried out using only a minimum set of features and a simple model of context. Experiments were performed on an Ikonos scene of a rural area in Germany. In our experiments, completeness and correctness values of 90% and better could be achieved, the CRF approach was clearly outperforming a standard Maximum-Likelihood-classification based on the same set of features.

1. INTRODUCTION

1.1 Motivation

The detection of settlement areas in satellite imagery is the basis for many applications, e.g. regional planning, the observation of urban expansion, or disaster prevention and management. In optical remote sensing images settlement areas have a heterogeneous appearance because they consist of a large number of different objects such as buildings, trees, and roads. The variety of these objects results in specific local patterns in the images. Whereas these patterns make a spectral classification of such areas very difficult, they can at the same time be exploited to improve the classification result if they are properly modelled. It is the main goal of this paper to model the contextual information contained in the local patterns of image features to improve the accuracy that can be achieved in the classification of settlement areas. In order to do so, we want to use Conditional Random Fields (CRF) (Kumar & Hebert, 2006) because of their ability to consider contextual relations between both the class labels and the observed image features of the image sites (i.e., pixels or segments). For this purpose, we will use radiometric and texture features from multispectral Ikonos data, i.e. from imagery having a resolution of 4 m. The parameters of the CRF will be learned from training data, and we will assess the effects of using the context information on the classification results.

1.2 Related Work

The methods that can be applied to detect settlement areas in satellite images depend on the resolution of these images. In images having a resolution better than about 2.5 m, a settlement is decomposed into buildings, roads, vegetation, and other objects. Various classification techniques have been proposed to extract these object classes, e.g. (Gamba et al., 2007). In images of 2.5 – 10 m resolution, which are our main interest here, the individual objects can no longer be discerned except for large structures. Buildings, roads, and urban vegetation are merged into a class 'settlement' which is characterized by a very heterogeneous distribution of the spectral components of the respective pixels. Hyperspectral data may help to overcome this problem (Herold et al., 2003), but the more common approach is to introduce textural features into classification, because they are better suited to characterize settlements, e.g. (Cheriyadat et al, 2007; Zhong & Wang, 2007). Various textural features have been used for urban classification, e.g. features based on the Grey-Level Co-occurrence Matrix (GLCM) (Smits & Annoni, 1999; Cheriyadat et al., 2007; Zhong & Wang, 2007), normalised grey-level histograms (Shackelford & Davis, 2003), or features related to the distribution of gradient orientation (Zhong & Wang, 2007).

These features can be used in any classification scheme. In a Bayesian statistical setting, the features of individual image sites are considered to be conditionally independent, which leads to a separate classification of each of the individual sites (Bishop, 2006). This approach has been found to lead to a salt-and-pepper-like appearance of the classification results. In order to improve the situation, context can be taken into account in the classification process. The simplest way of doing so is by post-processing the original classification results, taking into account the distribution of class labels in a local neighbourhood, e.g. (Gamba & Dell'Acqua, 2003). A more sophisticated approach uses statistical models of context. Among these, Markov Random Fields (MRF) (Besag, 1986) have found many applications in pattern recognition and remote sensing, e.g. (Tupin & Roux, 2005; Gamba et al., 2007). MRF can be used for representing texture, e.g. (Paget & Longstaff, 1998). In a Bayesian context, the main contribution of MRF is to act as a smoothness term on the class labels via a model for their local statistical dependencies (Besag, 1986; Kumar & Hebert, 2006).

* Corresponding author.

The features extracted from different sites are still assumed to be conditionally independent, and the interaction between neighbouring image sites is restricted to the class labels. Conditional Random Fields (Kumar & Hebert, 2006) were developed to overcome these restrictions. CRF provide a discriminative framework that can also model dependencies between the data and interactions between the labels and the data. In their experiments with man-made structure detection in natural terrestrial images, Kumar and Hebert (2006) could show that CRF outperform MRF.

Up to now, hardly any work has been done on classifying remotely sensed data using CRF. Zhong and Wang (2007) analyse images from Quickbird and SPOT with a multiple CRF ensemble model for the detection of settlement areas. They apply CRF to five groups of texture features and then fuse these results. The fusion process itself is based on a MRF taking into account the conditional probabilities provided by each of the CRF. Lu et al. (2009) use CRF on LiDAR data for simultaneously classifying the LiDAR data into terrain- and off-terrain-points and estimating a Digital Terrain Model from the off-terrain points. He et al. (2008) use CRF for building extraction from SAR data. Of these works, our new method is most closely related to (Zhong & Wang, 2007). However, our model is simpler because it only employs a single CRF that is applied to a feature vector taking into account radiometric and textural characteristics of the image. As the local dependencies of image data and class labels are modelled by a CRF in a very general way (Kumar & Hebert, 2006), we do not think it is necessary to use a MRF in order to fuse the output of a set of CRF. In our experiments, the effects of including a statistical model of context based on CRF on the classification results will be assessed by comparing the results of our new method to a standard maximum likelihood classification based on the same set of features. The main focus of this paper is on the benefits of using CRF for modelling context in classification and not on finding an optimum set of features for describing settlements.

2. MODELLING CONTEXT IN CLASSIFICATION USING CONDITIONAL RANDOM FIELDS

In many classification algorithms the decision for a class at a certain image site is just based on information derived at the regarded site, where a site might be a pixel, a square block of pixels in a regular grid or a segment of arbitrary shape. In fact, the class labels and also the data of neighbouring sites are often very similar or show characteristic patterns. Incorporating contextual information of neighbouring sites should improve the classification accuracy. The method described in this paper uses CRF for that purpose. In this section we want to give a brief overview on the CRF framework that is based on (Kumar & Hebert, 2006) and (Vishwanathan et al., 2006).

2.1 Conditional Random Fields (CRF)

The classification problem to be solved can be described as follows. We have observed image data $\mathbf{y}$. The image consists of image sites $i \in S$, where S is the set of all image sites. For each image site we want to determine its class x_i from a set of pre-defined classes. The class labels of all image sites can be combined in a vector $\mathbf{x}$ whose i^{th} component is the class of an individual image site i. Probabilistic classification methods determine the class labels so that they maximise the conditional probability $P(\mathbf{x} \mid \mathbf{y})$ of the class labels $\mathbf{x}$ given the observed data $\mathbf{y}$. CRF provide a discriminative framework for directly modelling $P(\mathbf{x} \mid \mathbf{y})$, which reduces the complexity of the involved models (Kumar & Hebert, 2006):

$$P(\mathbf{x}|\mathbf{y}) = \frac{1}{Z} exp\left(\sum_{i \in S} A_i(x_i, \mathbf{y}) + \sum_{i \in S} \sum_{j \in N_i} I_{ij}(x_i, x_j, \mathbf{y}) \right) \quad (1)$$

In Equation 1, $i \in S$ is the index of an individual image site, N_i is a certain neighbourhood of image site i, and thus j is an image site that is a neighbour to i. Z is a normalisation constant required to make $P(\mathbf{x} \mid \mathbf{y})$ a probability. The exact determination of Z is computationally intractable, which is the reason why approximate methods have to be used to determine the parameters of the model in Equation 1 and to maximise $P(\mathbf{x} \mid \mathbf{y})$ in the classification stage. In the exponent of Equation 1, the *association potential* A_i links the class label x_i of image site i to the data $\mathbf{y}$. Unlike with MRF, the association potential for an image site i may depend on the entire image $\mathbf{y}$. Thus, the data from neighbouring image sites are no longer considered to be conditionally independent. The second term in the exponent of Equation 1 is the *interaction potential* I_{ij}. It is responsible for modelling the dependencies between the labels x_i and x_j of neighbouring sites i and j and the data $\mathbf{y}$. This dependency of the interaction potential on the data is the second advantage of CRF over MRF. In MRF the interaction terms just depend on the labels, so that in many applications they only act as a kind of smoothness prior on the labels (Kumar & Hebert, 2006).

Any application of the CRF framework has to define what constitutes an image site and which classes are to be discerned. Furthermore, a model for the association and interaction potentials has to be found. We choose the image sites to be square blocks of pixels in a regular grid. The side length s of these squares is a parameter to be set by the user. We are only interested in a binary classification, so $x_i \in \{-1; 1\}$, where $x_i = 1$ means that image site i belongs to class *settlement* and $x_i = -1$ means that it belongs to the background. We model the CRF to be isotropic and homogeneous, hence the functions used for A_i and I_{ij} are independent of the location of image site i.

2.2 Association Potential

The association potential indicates how likely a site i is to belong to a label x_i given the observed data $\mathbf{y}$ and ignoring the other image sites. Kumar and Hebert (2006) suggest local discriminative classifiers for modelling the association potential by linking the association potential to the conditional probability $P'(x_i \mid \mathbf{y})$ of class x_i at image site i given the data $\mathbf{y}$:

$$A_i(x_i, \mathbf{y}) = log\, P'(x_i \mid \mathbf{y}) \quad (2)$$

The image data $\mathbf{y}$ are usually represented by image features that are determined from the original grey levels of the image. In order to put into practice the dependency of the association potential from the whole image, Kumar and Hebert (2006) define a site-wise feature vector $\mathbf{f}_i(\mathbf{y})$ which, though being computed specifically for site i, may depend on the entire image $\mathbf{y}$; usually the feature vector will be influenced by the data in a local neighbourhood that is not identical to the neighbourhood used for the interaction potential. Kumar and Hebert (2006) suggest using general linear models for $P'(x_i \mid \mathbf{y})$. For that purpose a feature space mapping $\mathbf{\Phi}(\mathbf{f})$ is required. It transforms the site-wise feature vectors $\mathbf{f}_i(\mathbf{y})$ into another feature space of

higher dimensions so that the decision surface becomes a hyperplane. Let $\mathbf{h}_i(\mathbf{y}) = \mathbf{\Phi}(\mathbf{f}_i(\mathbf{y}))$ be the site-wise transformed feature vector, with $\mathbf{\Phi}(\mathbf{f}_i(\mathbf{y})) = [1, \Phi_1(\mathbf{f}_i(\mathbf{y})), \ldots \Phi_N(\mathbf{f}_i(\mathbf{y}))]^T$ and Φ_k being arbitrary functions. The dimension of the transformed feature space is $N + 1$. In a generalised linear model, the conditional probability $P'(x_i \mid \mathbf{y})$ is described by Equation 3:

$$P'(x_i|\mathbf{y}) = \frac{1}{1+e^{-x_i \cdot \mathbf{w}^T \cdot \mathbf{h}_i(\mathbf{y})}} \quad (3)$$

where $\mathbf{w}$ is a vector of dimension $N + 1$. Its components describe the weights of the transformed features. These weights are the parameters of the association potential that have to be determined in a training phase. Fixing the first component of $\mathbf{h}_i(\mathbf{y})$ to 1 accommodates the bias parameter in the linear model in the exponent of Equation 3 (Bishop, 2006).

2.3 Interaction Potential

The interaction potential is a measure for the influence of the data $\mathbf{y}$ and the neighbouring labels x_j on the class x_i of site i. It can be linked to the conditional probability $P''(x_i = x_j \mid \mathbf{y})$ for the occurrence of identical labels at sites i and j given the data $\mathbf{y}$:

$$I_{ij}(x_i, x_j, \mathbf{y}) = \log P''(x_i = x_j | \mathbf{y}) \quad (4)$$

In the interaction potential, the data are represented by site-wise feature vectors $\boldsymbol{\psi}_i(\mathbf{y})$, which may have a different functional form than the vectors $\mathbf{f}_i(\mathbf{y})$ used for the association potential in order to accommodate features that are typical for neighbourhood dependencies. From the feature vectors $\boldsymbol{\psi}_i(\mathbf{y})$ and $\boldsymbol{\psi}_j(\mathbf{y})$ of two neighbouring sites a new vector of relational features $\boldsymbol{\mu}_{ij}(\mathbf{y}) = \boldsymbol{\mu}_{ij}(\boldsymbol{\psi}_i(\mathbf{y}), \boldsymbol{\psi}_j(\mathbf{y}))$ can be derived. Kumar and Hebert (2006) suggest concatenating the two vectors $\boldsymbol{\psi}_i(\mathbf{y})$ and $\boldsymbol{\psi}_j(\mathbf{y})$ or using some distance function. The interaction potential can be modelled as

$$I_{ij}(x_i, x_j, \mathbf{y}) = x_i x_j \mathbf{v}^T \boldsymbol{\mu}_{ij}(\mathbf{y}) \quad (5)$$

In Equation 5, the vector $\mathbf{v}$ contains the feature weights. They are the parameters of the model of the interaction potential and have to be determined by training. Kumar and Hebert (2006) give a geometric interpretation of the interaction potential: It partitions the space of the relational features $\boldsymbol{\mu}_{ij}(\mathbf{y})$ betweens the pairs that have the same class labels and pairs that have different labels. Thus, unlike with the well-known Ising model for MRF (Besag, 1986), it will moderate smoothing of neighbouring labels if there is a discontinuity of the features between the two sites.

We use $\boldsymbol{\psi}_i(\mathbf{y}) = \mathbf{f}_i(\mathbf{y})$, i.e. the features used for the interaction potential are identical to those used for the association potential. Furthermore, the component-wise absolute differences are used for the relational features $\boldsymbol{\mu}_{ij}$, i.e. $\boldsymbol{\mu}_{ij}(\mathbf{y}) = [1, |f_{i1}(\mathbf{y}) - f_{j1}(\mathbf{y})|, \ldots |f_{iR}(\mathbf{y}) - f_{jR}(\mathbf{y})|]^T$, where R is the dimension of the feature vectors $\mathbf{f}_i(\mathbf{y})$ and $f_{ik}(\mathbf{y})$ is the k^{th} component of $\mathbf{f}_i(\mathbf{y})$. The neighbourhood N_i of image site i consists of the four neighbouring image sites.

2.4 Parameter Learning and Classification

The parameters of the model for $P(\mathbf{x} \mid \mathbf{y})$ are the weights $\mathbf{w}$ and $\mathbf{v}$ of the association and interaction potentials, respectively. They can be combined to a parameter vector $\boldsymbol{\theta} = [\mathbf{w}^T, \mathbf{v}^T]^T$ that has to be estimated from training samples, i.e. a set $Y = \{\mathbf{y}_1, \ldots \mathbf{y}_M\}$ of M training images for which the class labels $X = \{\mathbf{x}_1, \ldots \mathbf{x}_M\}$ are known. If the parameters $\boldsymbol{\theta}$ are known, classification can be performed by maximising $P(\mathbf{x} \mid \mathbf{y})$ according to Equation 1. However, exact inference is computationally intractable for CRF (Kumar & Hebert, 2006). Vishwanathan et al. (2006) compare various methods for inference on CRF and come to the conclusion that Loopy-Belief-Propagation (LBP) (Frey & MacKay, 1998), which is a standard technique for performing probability propagation in graphs with cycles, provides the best results. It is thus used for classification in this work. In order to determine the parameters $\boldsymbol{\theta}$, $P(\mathbf{x} \mid \mathbf{y})$ is interpreted as $P(\mathbf{x} \mid \mathbf{y}, \boldsymbol{\theta})$, and $\boldsymbol{\theta}$ is estimated so that it maximises the conditional probability $P(\boldsymbol{\theta} \mid X, Y)$ or minimises the negative log-likelihood $L(\boldsymbol{\theta}) = -\log(P(\boldsymbol{\theta} \mid X, Y))$. An optimisation method that is frequently used is the BFGS Quasi-Newton method (Nocedal & Wright, 2006). If applied to minimise $L(\boldsymbol{\theta})$, it requires the computation of the gradients of $L(\boldsymbol{\theta})$, which in turn requires the selection of an approximate inference method (Vishwanathan et al., 2006). Following Vishwanathan et al. (2006), we use BFGS together with LBP for the simultaneous estimation of $\mathbf{w}$ and $\mathbf{v}$.

3. FEATURE EXTRACTION

In order to apply the CRF framework, the site-wise feature vectors $\mathbf{f}_i(\mathbf{y})$ that are used both for the association and the interaction potentials must be defined. It has to consist of appropriate features that can help to discriminate settlements from the background. In our application, we use two groups of features, namely gradient-based features $\mathbf{f}_{gi}(\mathbf{y})$ and colour-based features $\mathbf{f}_{ci}(\mathbf{y})$. Thus, the site-wise feature vector for site i contains both groups: $\mathbf{f}_i(\mathbf{y}) = [\boldsymbol{f}_{gi}(\mathbf{y})^T, \boldsymbol{f}_{ci}(\mathbf{y})^T]^T$. Both $\mathbf{f}_{gi}(\mathbf{y})$ and $\mathbf{f}_{ci}(\mathbf{y})$ contain features computed at two different scales λ_1 and λ_2. At scale λ_1, they are computed taking into account only the pixels inside the image site i (which is a square box of $s \times s$ pixels), whereas at scale λ_2 the pixels in a square of size $2 \cdot s$ centred at the centre of image site i are taken into account. Hence we do not only consider information derived at site i for the site-wise feature vectors $\mathbf{f}_i(\mathbf{y})$, but we also model dependencies between the image information of neighbouring sites. Of course, this principle could be expanded to a larger number of scales.

3.1 Features Based on Gradients

For determining the gradient-based features, we start by computing the gradient magnitude (Figure 1) and orientation for each pixel of the input image. All the gradient-based features are derived from a weighted histogram of the gradient orientations computed for each image site at both scales. Each histogram has 30 bins, so that each bin corresponds to an orientation interval of 6° width. Each bin contains the sum of the magnitudes of all gradients having an orientation that is within the interval corresponding to the bin. Summing over the magnitudes and not just counting the numbers of gradients falling into each bin is necessary to maintain the impact of strong magnitudes.

Three examples for histograms of different land cover types are shown in Figure 2. It shows that due to the heterogeneity of settlement areas, there are several strong peaks in this class, whereas cropland is nearly homogeneous and has a histogram showing low magnitudes. Thus, that mean *MG* and the variance *VG* of the histogram magnitudes are chosen as features to distinguish between textured and homogeneous areas. The third example in Figure 2 shows a road passing through cropland. In such a situation, the histogram shows only one strong peak as opposed to the settlement, where a larger diversity of orientations and thus a larger number of peaks can be observed. Thus, the number of bins *NG* with values above the mean was selected as the third gradient-based feature. All the features are normalised so that the values are in the interval [0, 1]. The gradient based feature vector $\mathbf{f}_{gi}(\mathbf{y})$ of image site *i* consists of six elements (three for each scale): $\mathbf{f}_{gi}(\mathbf{y}) = [MG_i^{(1)}, VG_i^{(1)}, NG_i^{(1)}, MG_i^{(2)}, VG_i^{(2)}, NG_i^{(2)}]^T$, where the upper index indicates the scale. We also tried to use the main orientation of the image site and the angle between the two largest peaks of the histogram as additional features. Neither modification resulted in any significant improvement of the classification performance.

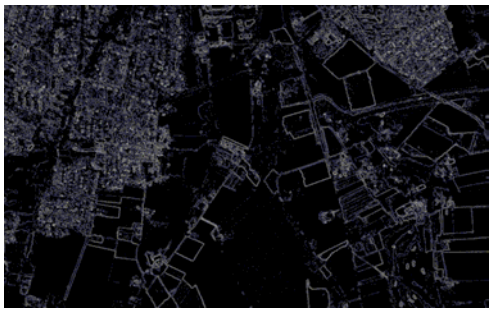


Figure 1. Gradient magnitude image of the test area.

3.2 Features Based on Colour

Figure 2 shows that in settlement areas we can expect a large variation of colours, whereas other land cover classes show a more homogeneous appearance. We carry out an IHS transformation and then proceed by analysing the hue image (Figure 3). For each image site *i* we compute the variance of the hue *VH* at both scales and normalise it so that its values are in the interval [0, 1]. The colour based feature vector of image site *i* has two components, namely *VH* for both scales: $\mathbf{f}_{ci}(\mathbf{y}) = [VH_i^{(1)}, VH_i^{(2)}]^T$. We also tried to use the mean hue as an additional feature, but it did not improve our results. We also tried to use other bands or combinations of bands, but using the hue band showed better performance than any other single band, and the consideration of other bands did not improve the results significantly while increasing the computational costs.

3.3 Feature Space Mapping

The site-wise feature vectors $\mathbf{f}_i(\mathbf{y})$ have a dimension of 8. As in (Kumar & Hebert, 2006), the transformed feature vectors $\mathbf{h}_i(\mathbf{y})$ are obtained by a quadratic expansion of the feature vectors $\mathbf{f}_i(\mathbf{y})$ so that the functions $\Phi_k(\mathbf{f}_i(\mathbf{y}))$ include all the $l = 8$ components of $\mathbf{f}_i(\mathbf{y})$, their squares and all their pairwise products. The dimension of the transformed feature vectors $\mathbf{h}_i(\mathbf{y})$ is $l + 1 + l \cdot (l + 1) / 2 = 45$. In case of the interaction potential, no feature space mapping is used. The dimension of the relational feature vectors $\boldsymbol{\mu}_{ij}(\mathbf{y})$ is 9. Using a feature space mapping for these relational feature vectors degraded the results in our tests, maybe because the feature space becomes too high-dimensional.

4. EXPERIMENTS

For our experiments we used the RGB bands of a multi-spectral Ikonos scene of a rural region near Herne, Germany. The resolution is 4 m. Two test areas having a similar type of land cover were cut out of the scene, each covering an area of 3.2 x 2.0 km^2. Ground truth was obtained by manually labelling these test areas on a pixel-level. In order for an area to be labelled as a settlement, it had to contain at least four houses; smaller groups of houses were ignored. One of the test areas and the related ground truth were used for training, whereas the other one served as our test scene. For the test scene, the ground truth could be used to evaluate the results.

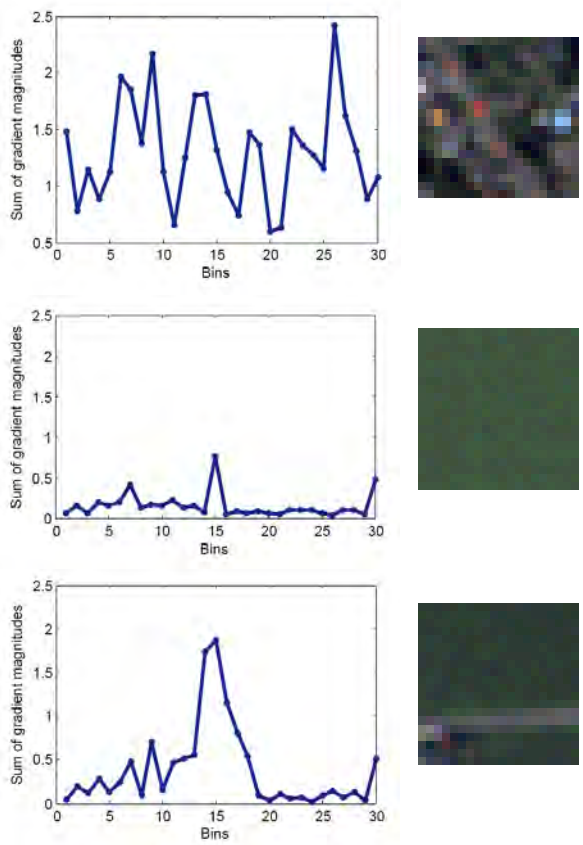


Figure 2. Gradient orientation histograms and the image patches they were computed from (s = 20 pixels). Upper row: settlement; centre: cropland; last row: cropland intersected by a road.

Figure 3. Hue image of the test area.

After having defined the size *s* of an image site, the features and the class labels were determined for all the image sites of the training area. An image site was labelled as belonging to class settlement if more than 50% of its pixels belonged to the

settlement class. The features and the class labels for the image sites of the training area were used to determine the parameters of the CRF. After that, the test scene was also subdivided into image sites of size *s*, the features were extracted for all image sites, and the parameters learned from the training data were used to determine the class of each image site by maximising $P(\mathbf{x} \mid \mathbf{y})$ using LBP. A reference classification was determined from the ground truth in the same way as the class labels for training were generated, i.e. by majority voting of the pixels in each image site. After that, completeness, correctness and quality (Heipke et al., 1997) were computed based on a comparison of the class labels of the image sites.

This procedure was applied using three different block sizes *s*, namely s = 4, 10, and 20 pixels, which resulted in 25000, 4000, and 1000 image sites, respectively. By using different block sizes, it should be possible to assess the influence of this parameter on the results. Furthermore, we carried out a standard Maximum-Likelihood (ML)-classification using $s = 4$ and $s = 10$ pixels and the same features as for the CRF, but only for the scale λ_1; using also λ_2 deteriorated the ML results. In the ML classification we used a normal distribution for the likelihood model $P(\mathbf{f}_i(\mathbf{y}_i) \mid x_i)$, determining the mean and covariance function from the training data. A comparison of the ML classification results and the results achieved by using CRF should highlight the influence of the statistical model of context.

The completeness, correctness, and quality achieved for the test scene in our experiments are shown in Table 4. The CRF-based method achieves completeness and correctness values of 90% and better in all cases except for s = 4, where completeness is slightly smaller. In comparison, the ML method also achieves 90% completeness, but correctness is very low (76%) for $s = 10$ pixels. For $s = 4$ pixels, the results are even worse. Using the CRF framework with its statistical model of context in the classification process significantly increases the quality of the results.

Method	s [pixel]	Completeness	Correctness	Quality
ML	4	77.6%	68.2%	57.0%
ML	10	90.7%	75.8%	70.3%
CRF	4	89.6%	90.3%	81.7%
CRF	10	92.9%	90.0%	84.2%
CRF	20	94.4%	91.6%	86.9%

Table 4. Evaluation of the classification results achieved for ML and for CRF using different block sizes *s*.

Figure 5 shows the ground truth and the results achieved both for CRF and ML classification for $s = 10$ pixels. The CRF results achieved for $s = 20$ pixels are shown in Figure 6. Examining these figures, it is obvious that the CRF approach tends to result in compact shapes. It works very well on the larger settlement areas. However, the smoothing effects of the context model cause small settlement areas to be missed. Small patches of non-settlement areas surrounded by settlement are also misclassified. These over-smoothing effects indicate that the impact of the interaction potential might be too strong. On the other hand, comparing the results of the CRF and ML classification results in Figure 5, the benefits of considering context become obvious. The ML results are much noisier. Large structures in settlements are not correctly detected, and there are many small false positives related to groups of trees. For the CRF method, there is a minor effect of the block size on the quality of the results: using $s = 20$ pixels, the completeness is 5% larger than for $s = 4$ pixels, because the features can be extracted more reliably if the block size is larger. However, a larger block size will reduce the level of detail of the results. Our experiments indicate that a value between s = 4 and s = 10 pixels might be optimal. Figure 7 shows a part of the test area for the CRF and ML classification using $s = 4$ pixels.

Figure 5. Test scene for $s = 10$ pixels. Class *settlement* is superimposed to the image in red. First row: ground truth; second row: CRF; third row: ML.

Figure 6. Results of CRF classification using $s = 20$ pixels. Class *settlement* is superimposed to the image in red.

Despite the in general somewhat poorer results of the CRF approach compared to larger block size, the shape of the settlement is well-preserved, whereas a reliable classification can not be achieved using the ML approach.

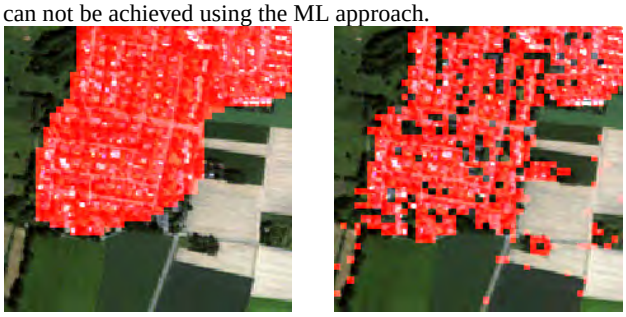


Figure 7. Section of the results of the Maximum-Likelihood-classification and the CRF-classification for $s = 4$.

Our results are quite promising, even more so because they were achieved using only a small set of features and a relatively simple model for the interaction potential. Using better features or a better context model could still improve the results.

5. CONCLUSION AND OUTLOOK

We have presented a new CRF-based approach for the classification of settlements in high resolution optical satellite imagery. CRFs allow incorporating contextual information into the classification process. The focus of this paper was on the impact of the context information on the classification results and not on a sophisticated selection of features. Tests on a multispectral Ikonos scene of 4 m resolution containing settlement areas of different size have shown that our CRF-based approach can achieve completeness and correctness values of over 90% for settlement areas and that it clearly outperforms ML classification based on the same set of features. Further research will focus on the extension of the framework to a classification of an arbitrary number of classes. The necessity of this already becomes obvious when trying to classify Ikonos panchromatic data of 1 m resolution with our approach. Settlements and forests are much harder to distinguish, which leads to unsatisfactory results. The situation could be improved by considering at least one more class, namely *forest*. Moreover the CRF framework should be applied to the results of a preliminary segmentation in order to obtain a more precise determination of the class boundaries. In this way, the problem of several classes existing in one site could also be reduced. Another goal for the future is an extension of the CRF framework to make it applicable to multi-temporal interpretation by considering spatial as well as temporal context, e.g. by introducing an additional temporal interaction potential.

REFERENCES

Besag, J. 1986. On the statistical analysis of dirty pictures. J. Royal Statistical Soc. Series B (Methodological) 48(3):259-302.

Bishop, C. M., 2006. Pattern recognition and machine learning. 1st edition, Springer New York, 738 pages.

Cheriyadat, A., Bright, E., Potere, D., Bhaduri, B., 2007. Mapping of settlements in high resolution satellite imagery using high performance computing. GeoJournal 69(1/2):119-129.

Frey, B. J. and MacKay, D. J., 1998. A revolution: belief propagation in graphs with cycles. In: Advances in Neural Information Processing Systems, 10, MIT Press, pp. 479-485.

Gamba, P., Dell'Acqua, F., 2003. Increased accuracy multiband urban classification using a neuro-fuzzy classifier. Int. J. Remote Sensing 24(4):827-834.

Gamba, P., Dell'Acqua, F., Lisini, G., Trianni, G., 2007. Improved VHR urban area mapping exploiting object boundaries. IEEE-TGARS 45(8):2676-2682.

He, W., Jäger, M., Reigber, A., Hellwich, O., 2008. Building extraction from polarimetric SAR data using mean shift and conditional random fields. In: Proc. of European Conference on Synthetic Aperture Radar, 3, pp. 439-442.

Heipke, C., Mayer, H., Wiedemann, C., Jamet, O., 1997. Evaluation of automatic road extraction. In: *IntArchPhRS* XXXII (3-4W2), pp. 151–160.

Herold, M., Gardner, M. E., Robert, D. A., 2003. Spectral resolution requirements for mapping urban areas. IEEE TGARS 41(9):1907-1919.

Kumar, S. and Hebert, M., 2006. Discriminative Random Fields. Int. J. Computer Vision 68(2) :179-201.

Lu, W., Murphy, K. P., Little, L. J., Sheffer, A., Fu, H., 2009. A hybrid conditional random field for estimating the underlying ground surface from airborne LiDAR data. IEEE TGARS 47(8):2913-2922.

Nocedal, J. and Wright, S. J., 2006. Numerical Optimization. 2nd edition, Springer New York, 664 pages.

Paget, R. and Longstaff, I. D., 1998. Texture synthesis via a noncausal nonparametric multiscale Markov Random Field. IEEE Transactions on Image Processing 7(6):925-931.

Shackelford, A. K. and Davis, C. H., 2003. A hierarchical fuzzy classification approach for high-resolution multispectral data over urban areas. IEEE TGARS 41(9):1920-1932.

Smits, P. C. and Annoni, A., 1999. Updating land-cover maps by using texture information from very high-resolution space-borne imagery. IEEE TGARS 37(3):1244-1254.

Tupin, F. and Roux, M., 2005. Markov Random Field on region adjacency graph for the fusion of SAR and optical data in radargrammetric applications. IEEE TGARS 43(8):1920-1928.

Vishwanathan, S., Schraudolph, N. N., Schmidt, M. W., Murphy, K. P., 2006. Accelerated training of conditional random fields with stochastic gradient methods. In: 23rd International Conference on Machine Learning, pp. 969-976.

Zhong, P. and Wang, R., 2007. A multiple conditional random fields ensemble model for urban area detection in remote sensing optical images. IEEE-TGARS 45(12):3978-3988.

ACKNOWLEDGEMENTS

The implementation of our method uses the CRF Toolbox for Matlab by K. Murphy & M. Schmidt (Vishwanathan et al., 2006): http://www.cs.ubc.ca/~murphyk/Software/CRF/crf.html

ESTIMATION OF TIMBER ASSORTMENTS USING LOW-DENSITY ALS DATA

M. Holopainen [a, *], M. Vastaranta [a], J. Rasinmäki [a], J. Kalliovirta [a], A. Mäkinen [a], R. Haapanen [b], T. Melkas [c], X. Yu [d], J. Hyyppä [d], H. Hyyppä [e]

[a] University of Helsinki, Department of Forest Resource Management, Finland -(markus.holopainen, mikko.vastaranta, jussi.rasinmäki, jouni.kalliovirta, antti.mäkinen)@helsinki.fi
[b] Haapanen Forest Consulting - reija.haapanen@haapanenforestconsulting.fi
[d] Metsäteho Ltd - timo.melkas@metsateho.fi
[c] Finnish Geodetic Institute - (xiaowei.yu, juha.hyyppa)@fgi.fi
[e] Helsinki University of Technology, Research Institute of Modelling and Measuring for the Built Environment, Finland - hannu.hyyppa@tkk.fi

KEY WORDS: Forestry, Inventory, Distributed, Laser scanning, Estimation

ABSTRACT:

The objective here was to analyse the effects of inventory errors on the prediction of assortment outturn volumes carried out in current airborne laser scanning (ALS) inventory method and forest-planning simulation computing in Finland. Harvested logging machine data of 12 clear-cutting stands (5300 trees) in Evo (southern Finland) study area was used as field reference of the study. Prediction error of assortment outturn volumes contains forest inventory, stem distribution generation, prediction of stem form and simulation of bucking errors. ALS inventory-related bias in estimated timber assortments ranged from -5.1 m^3/ha to 20.5 m^3/ha and RMSE from 6.0 m^3/ha to 46.2 m^3/ha. Accuracy of the estimated stem distributions varies in different stands. The results showed that the accuracy of the estimates of timber assortments is considerably poorer than the accuracy of stands mean characteristics.

1. INTRODUCTION

Standwise forest inventory (SWFI) and increasingly ALS inventory data acts as input data in forest management planning calculations. After the inventory, stand development and the effects of silvicultural treatments are simulated, using various models (e.g. Hynynen et al. 2002). In forest management computations, the quality of the input data describing the stand's present state has a decisive impact on the reliability of the output results (Haara 2005). The longer the reference period, the larger the output errors; thus, inaccurate input data are especially problematic in the case of forestry yield value determination throughout the rotation period. In addition, inaccurate input data cause significant nonoptimal losses in forest planning and forest silviculture if the timing of various treatments fails due to erroneous input data (e.g. Eid 2000, Eid et al. 2004, Holopainen and Talvitie 2006, Holopainen et al. 2009).

Airborne laser scanning (ALS) is the most accurate remote-sensing technique for standwise forest inventory providing accuracies (RMSEs) ranging between 10% and 27% for the mean volume at stand or plot level (e.g. Næsset 1997, 2002, Holmgren 2003, Lim et al. 2003, Packalén and Maltamo 2006, Holopainen et al. 2008). For comparison, the mean errors of traditional standwise field inventory (SWFI) used in operational forest management planning vary for mean volume from 16% to 38% in Finland (Poso 1983, Haara and Korhonen 2004, Saari and Kangas 2005). Current ALS data acquisition costs are comparable to those of SWFI. The two main approaches to deriving forest information from small-footprint ALS data have been those based on laser canopy height distribution (area-based method, Næsset 1997, 2002) and individual tree detection (Hyyppä and Inkinen 1999, Persson et al. 2002, Leckie et al. 2003, Popescu et al. 2003, Maltamo et al. 2004).

Acquisition of forest-planning data is currently in a phase of radical change. Several forest organizations in Finland are currently replacing traditional SWFIs with area-level ALS inventories in which low-density (less than two pulses per m^2) ALS data are used as an auxilliary data source. This new forest resource information provides new opportunities for forest management planning and e.g. forest estate valuation but, on the other hand, also sets new demands on forest development models and simulation methodologies. It, for example, offers several alternatives for forming stem distributions.

Area based ALS features can be utilized for the formation of stem diameter distributions in several ways. One alternative is to first estimate the mean stand characteristics and then apply stem distribution models based on theoretical distributions (e.g. the Weibull distribution). Another alternative is to use ALS features to directly estimate stem distribution parameters in a manner proposed and studied by Gobakken & Naesset (2004, 2005), Maltamo et al. (2006), Bollandsås and Naesset (2007) or by Breidenbach et al. (2008). Third possibility is to utilize stem distribution series measured for field plots used as reference in k-NN or k-MSN method (Packalén and Maltamo 2008).

Most of the ALS research in forest inventory has focused on the estimation of mean characteristics, such as plot or stand mean height or mean volume (e.g. Naesset 2002, Maltamo et al. 2006, Holopainen et al. 2008). However, from the standpoint of both forest value assessment and operative timber harvesting, the prediction of species-specific assortment outturn volumes, namely pulp wood and saw wood, is by far the most essential issue. For example, the economic value of a forest stand cannot be accurately determined on the basis of total stem volume only. Instead, information on tree species and the stem distribution is

* Corresponding author

required to reliably determine the distribution of the total stem volume in various assortments.

The significance of tree species-specific estimates in forest-planning simulation and optimization calculations is considerable (e.g. Holopainen et al. 2008). It is thus a great deficit, that the accuracy of tree species-specific estimates is considerably poorer than mean volume estimate. On the level of a forest stand (compartment), which generally is the unit of operations, Packalén and Maltamo (2007) obtained relative tree species-specific RMSEs from 28% (pine) to 62% (deciduous). These results are fully comparable to the traditional SWFI estimation, which is prone to error, as well. Haara and Korhonen (2004) investigated the accuracy of SWFI in eastern Finland. Their study showed that on the stand level the relative RMSE varied from 29% (pine) to 65% (deciduous), while the relative RMSE of stand mean total volume was 25%.

One problem in examination of timber assortment-level estimation accuracy of ALS inventory is that it requires sufficiently accurate ground reference data. In practice, the best method for acquiring such data is to use measuring data gathered by logging machines. However, to utilize these kinds of data as reference data for an ALS inventory, a rather complicated experimental arrangement is required to synchronize logging and imaging time schedules and identify felled trees.

The objective here was to analyse the effects of inventory errors on the prediction of assortment outturn volumes carried out in current ALS inventory method and forest-planning simulation computing in Finland. Harvested logging machine data was used as field reference of the study.

2. METHOD

2.1 Study area

The research material comprised of 12 clear-cut forest stands (Table 1) located in an approximately 2000-ha managed forested area in the vicinity of Evo, Finland (61.19° N, 25.11° E).

The compartments) were spruce-dominant (83%). The delineation of all study compartments was checked using global positioning system (GPS) measurements.

	Age	BA	N	Dg	Hg	Area
Mean	88	18.5	1228	30.5	19.1	1.1
Min	51	12.3	211	22.8	14.4	0.2
Max	123	24.0	9556	37.3	23.6	1.9
Stdev	22	3.1	2771	3.9	3.1	0.6

Table 1. Statistics of age (years), basal area (BA, m^2/ha), stem number (N, 1/ha), mean diameter (Dg, cm), mean height (Hg, m) and area (ha) of the compartments according to the standwise field invetory data.

2.2 Logging machine mensurations

Data obtained by the logging machines were utilized in the study as reference data. The logging machines gathered so called STM data according to the Standard for Forest Data and communication (StanForD 2006). An STM file includes data for each felled tree regarding the logging machine's position at the time of felling, stem diameters at 10-cm intervals from the felling height to the final bucking height, tree species, bucking parameters (e.q price matrix, demand matrix) and bucked timber assortment volumes. Bucked assortment volumes include volumes of saw- (minimum diameter>15 cm) and pulp wood (minimum diameter ≤ 7 cm).

The logging machine information obtained covered 5300 felled trees. An STM file was saved for each felled tree producing commercial timber. STM data were obtained for all trees felled in clear-cutting compartments. Stem distribution series, assortment outturn volumes and mean stock characteristics for each clear-cutting compartment were derived, using stem diameter and length information present in the STM files.

The diameter at breast-height (dbh) was derived as each stem's 12^{th} measured diameter (10-cm stump + 120 cm = 130 cm). Total tree height was estimated, based on the commercial timber height present in the STM file.

2.3 ALS inventory

The ALS data were acquired in midsummer 2006. The flying altitude was 1900 m. The density of the pulses returned within the field plots was 1.8/m^2 (only, first, intermediate or last; 1.3/m^2 if only or first pulses were considered). A digital elevation model (DEM) and consequently heights above ground level were computed by the data provider. Same-date aerial photographs were obtained with a Vexcel Ultracam digital camera, as well. The photographs were orthorectified, resampled to a pixel size of 0.5 m and mosaiced to a single image covering the entire area. The near-infrared (NIR), red (R) and green (G) bands were available.

Modelling field reference data for ALS-inventory were gathered from the study area in midsummer 2007. Treewise field measurements from 264 fixed-radius (10 m) plots were collected and plot level characteristics calculated. There was a 1-year gap between the acquisition of ALS data and field measurements and logging machine measurements; the latest growth was subtracted.

Several statistical and textural features were extracted from the ALS data and aerial photographs. The extraction window was 16 x 16 m, which has been used in operative ALS inventories in Finland. The features included means and standard deviations of spectral values of aerial photographs and ALS height and intensity, Haralick textural features (Haralick et al. 1973; Haralick 1979) derived from spectral values, ALS height and intensity, and standard texture referring to a set of averages and standard deviations of spectral values, ALS height and intensity. The height statistics for the first and last pulses were calculated as in Suvanto et al. (2005): mean and maximum height, standard deviation and coefficient of variation of height, heights at which certain relative amounts of laser points had accumulated as well as percentages of laser points accumulated at various relative heights. Only pulses exceeding a 2-m height limit were included to remove hits to ground vegetation and bushes. Finally, percentages of points under 2-m in height were added. The total number of features in the final dataset was 172. All features were standardized to a mean of 0 and std of 1. ALS-feature selection was based on the genetic algorithm method presented e.g. by Goldberg (1989). A reduced set of features (11 features, see Holopainen et al., 2008) was used in the estimation of stand characteristics.

The estimation method was k-NN, which has long been used in Finnish remote sensing -aided forest inventory applications (e.g. Kilkki and Päivinen 1987; Tokola 1990, Muinonen and Tokola 1990; Tomppo 1991). The nearest neighbours were determined by calculating the Euclidean distances between the observations in the n-dimensional feature space. The number of nearest neighbours was set to 5.

2.4 Determination of stem distributions

Based on mean characteristics obtained by the ALS inventory, stem distribution series were generated for each study compartment. Reference stem distribution series were derived for each compartment, using logging machine STM data. These series are based on the felled trees and do not include generation errors. Generation of stem distribution series was carried out with the SIMO system (Rasinmäki et al. 2009) incorporating the Weibull distribution for 1-cm diameter classes. Separate distribution models were applied for pine (Mykkänen 1986), spruce (Kilkki and Päivinen 1986) and birch (Siipilehto 1999).

2.5 Determination of assortment volumes

The stem form of each diameter class in the generated stem distribution series was predicted, using Laasasenaho's (1982) stem curves. The predicted stem form was then used to determine assortment volumes for each diameter class. Bucking was performed, using Näsberg's (1985) dynamic algorithm.

3. RESULTS

3.1 Accuracy in estimation of mean characteristics

ALS inventory results of basal area (BA), mean diameter (Dg) and mean height (Hg) of clear cut stands were compared to reference data calculated using logging machine STM data. Bias and RMSE values are shown in Table 2. Accuracies of the ALS inventory are at the same level as in previous studies in the same area (e.g. Holopainen et al. 2008).

		BA	Dg	Hg
ALS	Bias	3.5	-2.9	-0.6
ALS	Bias-%	17.8	-10.2	-2.8
ALS	RMSE	4.5	6	2.6
ALS	RMSE-%	22.7	21.2	12.4

Table 2. Accuracy of the ALS-inventory in the logging compartments, BA (m^2/ha), Dg (cm) and Hg (m).

3.2 Accuracy in estimation of timber assortments

Prediction error of assortment outturn volumes contains forest inventory, stem distribution generation, prediction of stem form and simulation of bucking errors. The combined effects of all error sources were examined in this study.

		Saw wood			Pulp wood		
		Pine	Spruce	Birch	Pine	Spruce	Birch
ALS	BIAS	8.9	20.5	-1.8	0.2	-5.1	14.6
ALS	BIAS%	100.8	19.4	-22.2	5.3	-21.1	123.9
ALS	RMSE	19.9	46.2	9.2	6.0	11.5	19.1
ALS	RMSE%	225.0	43.7	116.5	142.5	47.1	161.8

Table 3. Accuracy in estimation of timber assortments based on ALS, m^3/ha

Table 3 shows that ALS inventory-related bias ranged from -5.1 m^3/ha to 20.5 m^3/ha and RMSE from 6.0 m^3/ha to 46.2 m^3/ha. Based on these results, it should be noted that accuracy level ofmean characteristics provided by ALS inventory were not achieved for timber assortments. The great magnitude of the pine- and birch-related relative errors is a consequence of the relatively minor respective logging outturns. A more realistic view of the effect of inventory error on pine and birch assortment volumes can therefore be acquired by examining the absolute accuracy statistics.

3.3 Predicted stem distributions

True stem distribution series were determined by a logging machine (STM). In addition to the true stem distribution series, predicted stem distribution series were formed for each clear-cutting compartment investigated in the study. Predicted series were generated on the basis of mean stock characteristic output by the ALS inventory (ALS). The essential results concerning stem distribution series generation are presented in the stand level figures 1, 2 and 3.

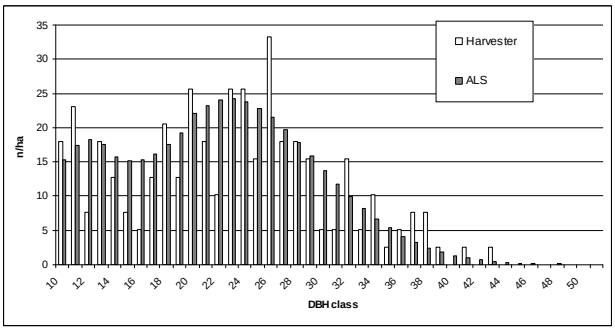


Figure 4. Predicted (ALS) and true stem distributions in a compartment with only slight estimation errors in basal area and mean diameter.

In figure 4, there is a compartment where basal area is overestimated by 1.3 m^2/ha and mean diameter underestimated by 0.8 cm. Still, the theoretical stem distribution does not represent accurately the true stem distribution series which inflicts errors varying from -3.9 m^3/ha to 16.9 m^3/ha in tree species specific saw log volumes and -10.1 m^3/ha to 6.0 m^3/ha to respective pulp wood volumes. Both variables basal area and mean diameter are overestimated, 5.3 m^2/ha and 2.1 cm, respectively, in the stand where stem distributions are presented in figure 5. Predicted stem distribution cannot describe the real variation between diameter classes. Errors in species specific saw log volumes varies from -5.5m^3/ha to 53.0m^3/ha and respective pulp wood volumes from -16.4m^3/ha to 5.3m^3/ha. Extreme-case is presented in figure 6. Mean diameter has been underestimated (5.2 cm) as basal area overestimated (2 m^2/ha). That kind of errors lead to totally biased stem distribution and inaccurate estimation of timber assortments. Errors in species specific saw log volumes varies from -91.0m^3/ha to 47.5m^3/ha

and respective pulp wood volumes from -8.3m^3/ha to 27.6m^3/ha.

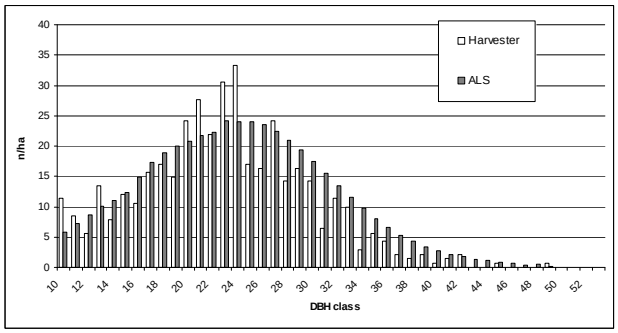


Figure 5. Predicted (ALS) and true stem distributions in a compartment where both basal area and mean diameter are overestimated.

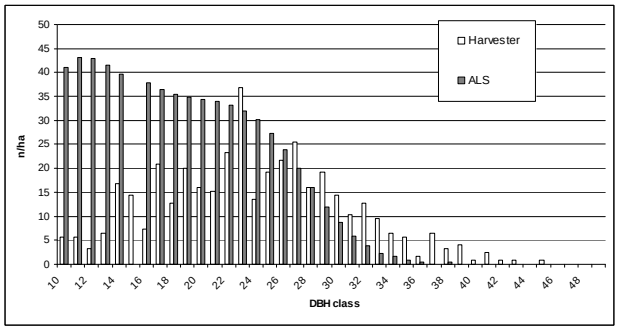


Figure 6. Predicted (ALS) and true stem distributions in a compartment, with overestimated basal area and underestimated mean diameter.

4. DISCUSSION

In the present study we investigated the accuracy of timber assortment volumes predicted by operational low-pulse, area-based, ALS inventory and forest management planning simulation in Finland. Analyses were performed for clear-cuttings, field reference data consisted of a total of 5300 stems felled in 12 logging compartments.

Errors in timber assortment level were higher than errors of clear cut stand's mean characteristics (see Tables 2 and 3). When analysing figures 1-3 it can be seen that quality of ALS-based diameter distribution series has great variance even if mean characteristics are close to true ones.

The most reliable results for the accuracy of the two inventory methods were derived for spruce saw timber and pulpwood assortments, because spruce was clearly the most common tree species in the clear-cutting compartments. In addition, one must take into account the fact that since the study data covered clear-cutting compartments only, the results cannot be generalized to consider all development classes.

Our results concerning the accuracy of low-pulse ALS inventory were slightly poorer than the plot-level ALS results in a study by Peuhkurinen et al. (2008) obtained with a k-NN method using ALS features and aerial photographs. Peuhkurinen et al. (2008) obtained relative RMSEs for spruce saw log volumes of 32.1% and bias of -2.3%, respectively. That accuracy is somewhat comparable with our results of uncertainty caused by ALS inventory in prediction of spruce saw log assortment.

Packalen & Maltamo (2008) tested low-pulse ALS data (0.7 pulses / m2) and spectral and textural features of calibrated aerial photograph in the prediction of species-specific diameter distributions. When they used a similar Weibull distribution method than we did, the accuracies of Scots pine logwood, Norway spruce logwood and deciduous logwood were (bias% in brackets) 40.97% (-3.27%), 61.06% (-25.78%), 142.93% (-37.62%).

In this study we wanted to focus in uncertainty of current forest management planning simulation methods in Finland, i.e. despite promising results of other possibilities to utilize area-level ALS inventory data (Gobacken & Næsset 2004, Maltamo et al. 2006, Packalén & Maltamo 2008), we used mean characteristics and theoretical weibull distribution series in the simulation.

REFERENCES

Bollandsås, O.M., and Næsset, E., 2007. Estimating percentilebased diameter distributions in uneven-sized Norway spruce stands using airborne laser scanner data, *Scand. J. For. Res*. 22, pp. 33–47.

Breidenbach, J., Gläser, C. and Schmidt, M., 2008. Estimation of diameter distributions by means of airborne laser scanner data. *Canadian Journal of Forest Research*, 38, pp. 1611-1620.

Eid, T. 2000. Use of uncertain inventory data in forestry scenario models and consequential incorrect harvest decisions. *Silva Fennica*, 34, pp 89-100.

Eid, T, Gobakken, T. and Næsset, E., 2004. Comparing stand inventories for large areas based on photo-interpretation and laser scanning by means of cost-plus-loss analyses. *Scandinavian Journal of Forest Research*, 19, pp. 512 – 523.

Gobakken, T., and Næsset, E., 2004. Estimation of diameter and basal area distributions in coniferous forest by means of airborne laser scanner data. *Scand. J. For. Res*., 19, pp. 529–542.

Gobakken, T. and Næsset, E., 2005.Weibull and percentile models for lidar-based estimation of basal area distribution. *Scand. J. For. Res*., 20, pp. 490-502.

Goldberg, D. E., 1989. Genetic algorithms in search, optimization, and machine learning. Addison-Wesley Publishing Company, Reading, Massachusetts, 412 p.

Haara, A. and Korhonen, K., 2004. Kuvioittaisen arvioinnin luotettavuus. *Metsätieteen aikakauskirja* 4, pp. 489-508. (in Finnish).

Haara, A., 2005. The uncertainty of forest management planning data in Finnish non-industrial private forestry. Doctoral Thesis. Dissertationes Forestales 8. University of Joensuu 34 p + 5 appendices.

Haralick, R. M., Shanmugan, K. and Dinstein, I., 1973. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics* 3, 6, pp. 610-621.

Haralick, R., 1979. Statistical and structural approaches to texture. *Proceedings of the IEEE 67*, 5, pp. 786-804.

Holmgren, J., 2003. Estimation of forest variables using airborne laser scanning. PhD Thesis. *Acta Universitatis Agriculturae Sueciae, Silvestria* 278, Swedish University of Agricultural Sciences, Umeå, Sweden.

Holopainen, M. and Talvitie, T., 2006. Effects of data acquisition accuracy on timing of stand harvests and expected net present value. Silva Fennica 40, 3, pp. 531-543.

Holopainen, M., Haapanen, R., Tuominen, S. and Viitala, R. 2008. Performance of airborne laser scanning- and aerial photograph-based statistical and textural features in forest variable estimation. In Hill, R., Rossette, J. and Suárez, J. 2008. *Silvilaser 2008 proceedings*, pp. 105-112.

Holopainen, M., Mäkinen, A., Rasinmäki, J., Hyyppä, J., Hyyppä, H., Kaartinen, H., Viitala, R., Vastaranta, M. and Kangas, A., 2009. Effect of tree level airborne laser scanning accuracy on the timing and expected value of harvest decisions. *European Journal of Forest Research*, in press.

Hynynen, J., Ojansuu, R., Hökkä, H., Siipilehto, J., Salminen, H. and Haapala, P., 2002. Models for predicting stand development in MELA System, Finnish Forest Res. Inst. Res. Pap. 835, pp. 1–116.

Hyyppä, J. and Inkinen, M., 1999. Detecting and estimating attributes for single trees using laser scanner. *The Photogrammetric Journal of Finland*, 16, pp. 27-42.

Kilkki, P. and Päivinen, R., 1986. Weibull function in the estimation of the basal-area DBH-distribution. *Silva Fenn*. 20, pp. 149–156.

Kilkki, P. and Päivinen, R., 1987. Reference sample plots to combine field measurements and satellite data in forest inventory. Department of Forest Mensuration and Management, University of Helsinki. Research notes, 19, pp. 210-215.

Laasasenaho, J., 1982. Taper curve and volume functions for pine, spruce and birch. *Communicationes. Institute Forestalis Fenniae* 108. 74 p.

Leckie, D., Gougeon, F., Hill, D., Quinn, R., Armstrong, L. and Shreenan, R., 2003. Combined high-density lidar and multispectral imagery for individual tree crown analysis. *Can. J. For. Res.* 29, pp. 633–649.

Lim, K., Treitz, P., Wulder, M., St.Onge, B. and Flood, M., 2003. LIDAR remote sensing of forest structure. *Progress in Physical Geography* 27, pp. 88-106.

Maltamo, M., Eerikäinen, K., Pitkänen, J., Hyyppä, J. and Vehmas, M., 2004. Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions. *Remote Sens. Environ*. 90, pp. 319–330.

Maltamo, M., Malinen, J., Packalén, P., Suvanto, A. and Kangas, J., 2006. Nonparametric estimation of stem volume using airborne laser scanning, aerial photography, and stand-register data. *Can J For Res* 36, pp. 426-436.

Muinonen, E. and Tokola, T. 1990. An application of remote sensing for communal forest inventory. Proceedings from SNS/IUFRO workshop: The usability of remote sensing for forest inventory and planning, 26-28 February 1990, Umeå, Sweden. Remote Sensing Laboratory, Swedish University of Agricultural Sciences, Report 4, pp. 35-42.

Mykkänen, R., 1986. Weibull-funktion käyttö puuston läpimittajakauman estimoinnissa. M. Sc. thesis. University of Joensuu, Faculty of Forestry. 80 p. (In Finnish).

Næsset, E., 1997. Estimating timber volume of forest stands using airborne laser scanner data. *Remote Sens. Environ*. 61, pp. 246-253.

Naesset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sens. Environ*. 80, pp. 88-99.

Näsberg, M., 1985. Mathematical programming models for optimal log bucking. Linköping:Department of Mathematics, Linköping University, 1985. Linköping studies in science andtechnology. Dissertation 132. 200 pages.

Packalén, P. and Maltamo, M., 2006. Predicting the plot volume by tree species using airborne laser scanning and aerial photographs. *Forest Science*, 56, pp. 611-622.

Packalén, P. and Maltamo, M., 2007. The k-MSN method in the prediction of species specific stand attributes using airborne laser scanning and aerial photographs. *Remote Sensing of Environment*, 109, pp. 328-341.

Packalén, P. and Maltamo, M., 2008. Estimation of species-specific diameter distributions using airborne laser scanning and aerial photographs. *Canadian Journal of Forest Research* 38, pp. 1750–1760.

Persson, Å., Holmgren, J. and Söderman, U., 2002. Detecting and measuring individual trees using an airborne laser scanner. *Photogrammetric Engineering and Remote Sensing* 68, pp. 925-932.

Peuhkurinen, J., Maltamo, M. and Malinen, J., 2008. Estimating species-specific diameter, distributions and saw log recoveries of boreal forests from airborne laser scanning data and aerial photographs: a distribution-based approach. *Silva Fennica*, 42, pp. 625-641.

Popescu, S., Wynne, R., Nelson, R., 2003. Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. *Canadian Journal of Forest Research* 29, pp. 564–577.

Poso, S., 1983. Basic features of inventory by compartments. *Silva Fennica* 17, pp. 313–349 (in Finnish).

Rasinmäki, J,, Kalliovirta, J, and Mäkinen, A, 2009. SIMO: an adaptable simulation framework for multiscale forest resource data. Comput Electron Agric 66, pp.76-84.

Saari, A. and Kangas, A., 2005. Kuvioittaisen arvioinnin harhan muodostuminen. *Metsätieteen aikakauskirja* 1, 2005, pp. 5-18.

StanForD, 2006. Skogsforsk: Standard for forest data and communications. Available online at www.skogsforsk.se/.

Suvanto, A., Maltamo, M., Packalén, P. and Kangas, J., 2005. Kuviokohtaisten puustotunnusten ennustaminen laserkeilauksella. *Metsätieteen aikakauskirja*, 2005, pp. 413-428.

Tokola, T., 1990. Satelliittikuvan ja VMI-koealatiedon käyttö metsätalousalueen puuston inventoinnissa. Joensuun yliopisto, metsätieteellinen tiedekunta. Lisensiaattitutkimus. 53 p.

Tomppo, E., 1991. Satellite image-based national forest inventory of Finland. *International Archives of Photogrammetry and Remote Sensing*, 28, pp. 419-424.

TOWARDS FULLY AUTOMATIC PHOTOGRAMMETRIC RECONSTRUCTION USING DIGITAL IMAGES TAKEN FROM UAVS

A. Irschara[*,a], V. Kaufmann[b], M. Klopschitz[a], H. Bischof[a], F. Leberl[a]

[a] Institute for Computer Graphics and Vision, Graz University of Technology, Inffeldgasse 16, A-8010 Graz, Austria – {irschara,klopschitz,bischof,leberl}@icg.tugraz.at
[b]Institute of Remote Sensing and Photogrammetry, Graz University of Technology, Steyrergasse 30, A-8010 Graz, Austria – viktor.kaufmann@tugraz.at

KEY WORDS: Vision, Robotics, Reconstruction, Matching, Automation, Accuracy

ABSTRACT:

We argue that the future of remote sensing will see a diversification of sensors and sensor platforms. We argue further that remote sensing will also benefit from recent advances in computing technology to employ new algorithms previously too complex to apply. In this paper we support this argument by three demonstrations. First, we show that an unmanned aerial vehicle (UAV) equipped with digital cameras can provide valuable visual information about the Earth's surface rapidly and at low cost from nearly any viewpoint. Second, we demonstrate an end-to-end workflow to process a sizeable block of such imagery in a fully automated manner. Thirdly, we build this workflow on a novel computing system taking advantage of the invention of the Graphics Processing Unit (GPU) that is capable of performing complex algorithms in an acceptable elapsed time. The transition to diverse imaging sensors and platforms results in a requirement to deal with unordered sets of images, such as typically collected from a UAV, and to match and orientate these images automatically. Our approach is fully automated and capable of addressing large datasets in reasonable time and at low costs on a standard desktop PC. We compare our method to a semi-automatic orientation approach based on the PhotoModeler software and demonstrate superior performance in terms of automation, accuracy and processing time.

1. INTRODUCTION

Aerial photography has been the workhorse of remote sensing. Satellite imagery has augmented the remote sensing tool box since the launch of Landsat in 1972. Both aerial and satellite imaging result in very ordered and industrially planned image datasets. Recently, however, one can see a diversification of the image inputs for remote sensing (Eissenbeiss et al., 2009). Photography from handheld amateur cameras, from balloons and unmanned aerial vehicles (UAVs), all are subject to intensive research into their applicability to tasks previously reserved to industrial solutions. In the last few years, advances in material science and control engineering have turned unmanned aerial vehicles into cost efficient, flexible and rapidly deployable geodata acquisition platforms. For instance the micro-drone md4-200 (http://www.microdrones.com) depicted in Figure 1 has the ability for vertical take off and landing, provides position hold and autonomous way-point navigation and is equipped with a standard digital consumer camera that can be tilted (up to 90°) to capture images from different angles. Thus, a UAV can act as a virtual eye in the sky capable to provide visual information about an object which otherwise cannot be obtained. Therefore, photogrammetric reconstruction based on imagery taken from UAV systems is of high interest and has been addressed by many authors, e.g. in the context of digital surface model (DSM) extraction (Förstner and Steffen, 2007), archaeological preservation (Scaioni et al., 2009) and agricultural survey (Grenzdörffer et al., 2008). According to (Colomina et al., 2008), UAVs are a new paradigm for high-resolution low-cost photogrammetry and remote sensing, especially given the fact that consumer grade digital cameras provide a sufficiently high accuracy for many photogrammetric tasks (Gruen and Akca, 2008). The presence of on board navigation, Global Positioning System (GPS) and Inertial Measurement Units (IMUs) allows UAVs to act as autonomous systems that fly in the air and sense the environment. Due to the low operation altitude, UAVs achieve a very high resolution

Figure 1: Micro-drone md4-200 with attached PENTAX Optio A40.

in terms of ground sampling distance and can therefore compete with airborne large format digital camera system (e.g. UltraCamXp (http://www.microsoft.com/ultracam)).

Although, recent UAVs are most often equipped with GPS/INS positioning systems and orientation sensors, the output of these sensors does in general not achieve the required accuracy to provide direct georeferencing of the acquired imagery (Eugster and Nebiker, 2009). Hence, image based methods, referred to as structure from motion in the computer vision literature (Hartley and Zisserman, 2000), are necessary techniques to determine the exterior camera orientations. There exists a variety of approaches that address the 3D reconstruction problem from videos and ordered sets of still images, e.g. (Pollefeys et al., 2004). Real time performance for camera motion recovery on modest hardware is reported (Nistér et al., 2004), but working incrementally on a frame by frame basis leads to the inherent problem of error accumulation and drift (Steffen and Förstner, 2008). Furthermore, sequential processing is only possible for very ordered, industrially planned image datasets, such as manned airborne and spaceborne remote sensing imagery. The transition to diverse imaging sensors and platforms results in a requirement to deal with unordered sets of images. This is especially true for images captured by highly maneuverable UAV systems that allow random flight paths, hence deliver unordered image datasets.

Therefore, in practice, wide baseline matching methods that are

* Corresponding author.

able to establish geometric relations between images, which are (widely) separated in time/space, are necessary in order to obtain consistent 3D models. These methods have been shown to even work on very uncontrolled image collections such as images from the web (Snavely et al., 2006), but require a high degree of computational effort. Recently, (Agarwal et al., 2009) presented a distributed computing engine based on a cluster of 500 computing cores to automatically reconstruct 3D scenes from large image collections. Our system shares algorithmic similarities with their approach, but in contrast to rely on hundreds of computer clusters, we leverage the parallel computing power of current GPUs to accelerate several processing steps. We follow the concept of General-Purpose computing on Graphic Processing Units (GPGPU) and use Nvidia's Compute Unified Device Architecture (CUDA) toolchain for our implementation. Our proposed approach is fully automated and capable of addressing large datasets in reasonable time and at low costs on a standard desktop PC.

2. UAVS AS PHOTOGRAMMETRIC SENSOR PLATFORMS

The main advantage of a UAV system acting as a photogrammetric sensor platform over more traditional manned airborne or terrestrial surveys, is the high flexibility that allows image acquisition from unconventional viewpoints. Consider Figure 2: While the camera network in standard airborne and terrestrial surveys is normally restricted to flight lines or street paths, a UAV system enables more flexible, e.g. turntable like network configurations, that maximize scene coverage and allow superior accuracy in terms of triangulation angles. Furthermore, the photogrammetric network planning task (Chen et al., 2008) can be optimized and adapted to the scene since nearly any desired viewpoint can be reached. Moreover, networks of multiple, synchronously flying UAVs (Quaritsch et al., 2008) could be utilized to deliver multi-view information simultaneously, which opens the possibility to reconstruct also non-rigid objects over time.

The remainder of the paper is organized as follows. In the next Section we describe in detail our structure from motion system which is able to operate on unordered datasets, such as typical images captured by a UAV system. In Section 4. we show results of our method and compare our system to a standard semi-automatic approach based on the PhotoModeler software. Finally, Section 5. concludes our work.

3. 3D RECONSTRUCTION SYSTEM

Our 3D reconstruction system is able to automatically match unordered sets of images and to determine the exterior camera orientations and sparse tie points without prior knowledge of the scene. The system mainly consists of three processing steps: Feature extraction, matching and finally structure from motion computation. Figure 3 gives an overview of our reconstruction pipeline. A prerequisite of our system is that the intrinsic camera parameters are known and constant. We use the calibration method described in (Irschara et al., 2007) to simultaneously estimate the focal length, principal point and radial distortion parameters, standard values are assumed for the remaining intrinsics (i.e. zero skew and unit aspect ratio).

In general, calibrated camera settings are not strictly necessary for Euclidean 3D modeling, since self-calibration methods (Pollefeys et al., 1999) exist, but robustness and accuracy is normally greatly improved for image collections with known intrinsics. Furthermore, also an increase in processing speed is achieved due to the lower dimensionality of the problem.

http://www.nvidia.com

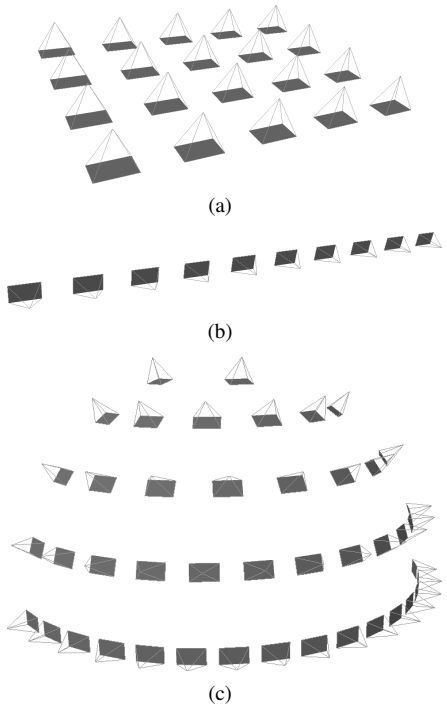


Figure 2: Typical camera networks used for aerial (a) and (b) terrestrial survey. In general, a UAV system allows the acquisition of more flexible photogrammetric camera networks, like the configuration depicted in (c), that enables a regular sampling of the visual hull of the scene of interest.

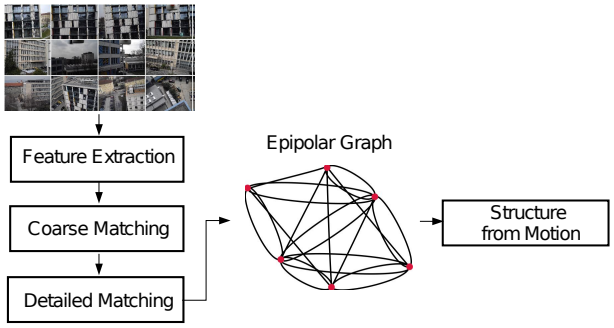


Figure 3: Overview of the main processing steps of our reconstruction pipeline.

3.1 Feature Extraction

Our system utilizes the very effective SIFT keypoint detector and descriptor (Lowe, 2004) to represent point features. SIFT features are invariant to scale and rotation and partially invariant to viewpoint and illumination changes. Hence, these kind of features are very suitable for wide baseline matching and have been found to be highly distinctive and repeatable in performance evaluation (Mikolajczyk et al., 2005). In particular we rely on the publicly available SiftGPU software. On recent GPUs, a speedup exceeding twenty over a single core CPU implementation is reached.

3.2 Matching

Unlike feature point tracking in video sequences, where correspondence search can be restricted to local regions, matching of

http://cs.unc.edu/~ccwu/siftgpu

unordered still images essentially requires exhaustive search between all image pairs and all features seen therein. Hence, the matching costs are quadratic in the total number of extracted features from the image database. Note, the number of SIFT features from a medium sized image (e.g. 4000×3000 pixel) normally exceeds a value of 10000. For a small image database consisting of 1000 images, more than 10 million SIFT keys are detected, this translates into 100 billion descriptor comparisons that are necessary for exhaustive nearest neighbor search. This is a considerable amount of computation, which turns out to be a prohibitively expensive operation executed on a single CPU.

To make the correspondence search more tractable, we divide the matching procedure into two submodules. First, we build upon work on efficient image retrieval (Nistér and Stewenius, 2006) and use a vocabulary tree to determine an image-to-image similarity score. Second, we take advantage of the high computational power of modern GPUs to establish putative correspondences between the feature sets of relevant image pairs.

3.2.1 Coarse Matching Inspired by recent advantages in image search, we use a vocabulary tree approach and inverted file voting (Sivic and Zisserman, 2003) for coarse matching of potentially similar images. The vocabulary tree based database representation is very efficient in terms of memory usage and allows an extremely fast determination (in the order of some milliseconds) whether two images are similar or dissimilar. Hence, by considering only the most relevant candidate images for pair-wise matching, the computational effort can be reduced significantly.

The vocabulary tree is constructed by offline training using hierarchical k-means clustering of millions of SIFT features (extracted from a generic image database) and gives a quantized approximation of the high dimensional descriptor space. Since k-means clustering of large datasets is a time consuming operation, we employ a CUDA based approach executed on the GPU to speed up clustering.

The vocabulary tree concept relies on the following basic assumption: if the similarity between two features $sim(f_i, f_j)$ is high, then there is a relatively high probability that the two features are assigned to the same visual word $w(f_i) \equiv w(f_j)$, i.e. the features reach the same leaf node in the vocabulary tree. Based on the quantized features from a query image $\mathcal{Q}$ and each database image $\mathcal{D}$ a scoring of relevance is derived. Typical scoring functions are based on a vector model, as for instance the *tf-idf* (term frequency, inverse document frequency), which delivers a relative document ranking according to the degree of similarity to the query. In contrast to that, in our system we rely on a scoring function that gives an absolute score of similarity based on a probabilistic model (Singhal, 2001, Irschara et al., 2009). This model allows a direct determination whether a document image is likely to match a query image.

3.2.2 Pairwise Feature Matching A variety of approaches have been proposed to speedup nearest neighbor matching in high-dimensional spaces (like the 128-dimensional SIFT descriptor space). Among the most promising methods are randomized kd-trees (Anan and Hartley, 2008) with priority search and hierarchical k-means trees (Fukunaga and Narendra, 1975). These algorithms are in general designed to run on a single CPU and are known to provide speedups of about one or two orders of magnitude over linear search, but the speedup comes with the cost of a potential loss in accuracy (Muja and Lowe, 2009). On the other hand, given that the number of features is limited to some thousands, nearest neighbor search, implemented as a dense matrix multiplication on recent graphics hardware, can achieve an equivalent speedup, but delivers the exact solution. Hence, we employ a GPU accelerated feature matching approach based on the CUBLAS library.

http://www.vis.uky.edu/ stewe/ukbench/

3.3 Epipolar Graph

After matching relevant images to each query view, geometric verification, based on the Five-Point algorithm (Nistér, 2004) is performed. Since matches that arise from descriptor comparisons are often highly contaminated by outliers, we employ a RANSAC (Fischler and Bolles, 1981) algorithm for robust estimation. In its basic implementation, RANSAC acts as hypothesize-and-verify approach. In the same spirit as (Nistér, 2005) we explicitly divide the RANSAC algorithm into two steps. First, we generate all our N relative pose hypotheses with a minimal number of five points. Second, we score the hypotheses based on the truncated Sampson error (Hartley and Zisserman, 2000) against each other. Note, the scoring procedure can be easily parallelized, hence we employ a CUDA based scoring approach in our reconstruction system.

In order to decide whether two images satisfy an epipolar geometry, we compute the RANSAC termination confidence,

$$p = 1 - \exp(N \log(1 - (1 - \epsilon)^s)) \quad (1)$$

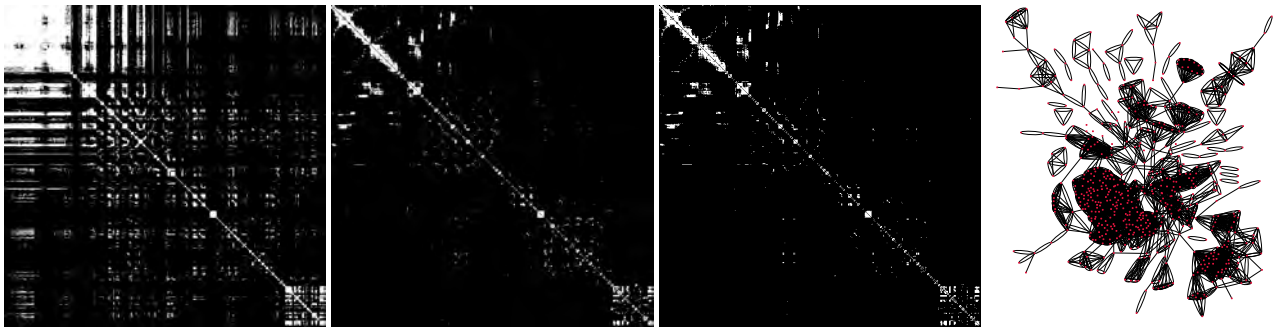
where N is the number of evaluated models, $w = 1-\epsilon$ the probability that any selected data point is an inlier, and $s = 5$ is the cardinality of the sample point set used to compute a minimal model. We require $p > 0.999$ in order to accept an epipolar geometric relation. In our experiments, we used up to $N = 2000$ models which corresponds to a maximal outlier fraction of $\epsilon = 0.67$. The epipolar graph of the UAV-dataset is shown in Figure 4(d).

3.4 Structure from Motion

Our structure from motion approach follows a greedy strategy, similar to the one described in (Irschara et al., 2007). Starting from a reliable image triplet, new views are incrementally registered by robust camera resectioning based on the Three-Point algorithm (Haralick et al., 1991) inside a RANSAC loop. Incremental Euclidean bundle adjustment is used to simultaneously refine structure (3D points) and motion (camera matrices).

4. RESULTS AND DISCUSSION

In our experiments we performed two test-flights with the micro-drone md4-200 and an attached PENTAX Optio A40 camera as depicted in Figure 1. The camera was precalibrated and the zoom was fixed to a wide angle setting. The survey was performed by manual remote control, 615 still images with a resolution of 4000×3000 square pixels were captured from different viewpoints. Furthermore, eight ground control points were determined using a total station (with an accuracy of $\epsilon \pm 1cm$, see Figure 8). This data is considered as ground truth and is later used to asses the object space error of the automatic computed structure from motion results. Figure 4 shows the affinity matrix according to the probabilistic scoring used for coarse matching. On average each image is only matched with 84 potentially similar views, which gives a speedup of approximately seven compared to a full exhaustive search. Still, 86% of potential epipolar relations are found. Note, the average degree of image overlap in this dataset is relatively high. A much higher speedup would be achieved if one considers larger datasets with a sparser image overlap. Since the epipolar graph of the UAV-datasets is not fully connected (see Figure 4(d)), several individual 3D reconstructions are obtained.



(a) (b) (c) (d)

Figure 4: (a) Image affinity matrix according to the probabilistic model of relevance. (b) Epipolar adjacency matrix computed by exhaustive image matching and geometric verification. A white entry in the matrix indicates that an epipolar geometry between two images I_i and I_j could be computed. (c) Successfully recovered epipolar geometries by considering only relevant images according to (a). (d) Epipolar connectivity graph of the whole dataset, clusters in the graph represent a high degree of geometric connectivity.

	CPU [s]	GPU [s]
SIFT (4000×3000 pixel)	10	0.4
Coarse Matching	0.5	0.05
Matching (5000×5000)	$k \times 1.1$	$k \times 0.044$
RANSAC-H (5-pt, N=2000)	$k \times 0.1$	-
RANSAC-V ($\|C\|$=5000, N=2000)	$k \times 0.12$	$k \times 0.02$
Structure from Motion [h]	1	-
Total Time [h] (615 views, $k = 84$)	21	3.5

Table 5: Comparison of processing timings between execution on a single core CPU (Intel Pentium D 3.2Ghz) vs. a GPU accelerated implementation (Nvidia GeForce GTX280). RANSAC-H stands for the hypotheses generation step based on the Five Point algorithm, RANSAC-V for the evaluation module. N is the maximal number of hypothesis, $|C|$ the number of putative correspondences used for evaluation, and k reflects the number of considered images for detailed feature matching and geometric verification.

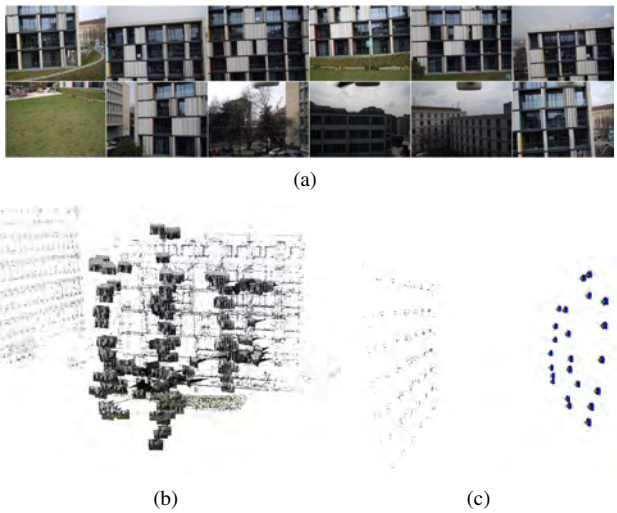
Figures 6 and 9 show visual results of the two largest connected reconstruction results, denoted as R1 (239 registered images) and R2 (68 registered images) through our experimental evaluation.

Table 5 gives typical processing times of the modules involved in our system and compares timings of a single CPU execution with timings achieved with GPGPU support. Regarding feature extraction and matching, the speedup induced by the GPU is about one order of magnitude.

4.1 Accuracy Analysis

We compare our fully automatic structure from motion approach to the semi-automatic PhotoModeler software (version 6) for the task of exterior image orientation. Since it turns out that processing 615 images is impracticable for a semi-automatic system, we restrict our evaluation to a subset of 23 manually selected images from one building facade (corresponding to result R1, see Figure 6). The processing steps of the PhotoModeler approach include the semi-automatic measurement of tie and control points, bundle adjustment and fine tuning. Four different orientation methods were conducted: selfcalibration with constant/variable intrinsics and with/without reference point constraints by using fifteen 3D control points, respectively. All methods give consistent results, on average a reprojection error of 0.5 pixel is reported. A detailed, quantitative comparison of the PhotoModeler

http://www.photomodeler.com



(a) (b) (c)

Figure 6: Orientation result R1: (a) Sample input images and (b) perspective view of camera orientations (239) and respective 3D points (58791) obtained by our automatic structure from motion system. (c) Orientation result obtained by semi-automatic processing using the PhotoModeler software, a subset of 23 manually selected images is used.

orientation output with results from our structure from motion pipeline is summarized in Table 7.

The semi-automatic approach, based on the PhotoModeler software, was performed by an expert user, the orientation of a subset of 23 images still requires about eight man hours (and is troublesome and strenuous work). On the other hand, with our fully automated system, all 615 images can be processed at once and within a timeframe of 3.5 hours on a standard PC and a single GPU. We achieve identical results in terms of reprojection error, but with a higher confidence in the solution, since many more tie points are utilized. Furthermore, the automatic approach is scalable and allows registration of many more images much faster. For instance, in our pipeline processing one image takes about 20s, whereas orientation with the PhotoModeler software requires more than 20min man workload.

4.1.1 Object Space Error The reprojection error is a suitable measure to assess the precision of camera orientations in image space, but for a practical application, the error in object space is of interest. Therefore, we rely on control points measured by a

	PhotoModeler	sfm-approach
# processed views	**23**	**615**
# registered views (R1)	23	239
# 3D points	237	58791
avg. # points/image	**99**	**3160**
avg. # rays/3D point	10	13
avg. triangulation angle	10°	6.7°
avg. reprojection error	0.458	0.460
processing time [h]	8	3.5
processing time/image [s]	**1252**	**20**

Table 7: Comparison of the semi-automatic PhotoModeler orientation to our proposed fully-automatic structure from motion system (sfm-approach), the values correspond to reconstruction result R1 (see Figure 6).

total stations to estimate an absolute error measure. The landmarks are determined at well localized structures, like building corners and junctions (see Figure 8). Thus, image measurements with respect to the corresponding landmark are easily to establish. For each image we estimate the 2D coordinates belonging to the 3D control point (manually by visual inspection) and link the measurements into point tracks. In practice, we only use a subset of images to measure observations, but ensure that for each control point at least three measurements are provided and the triangulation angle is sufficiently high ($\bar{\alpha} > 20°$). Next, we use a linear triangulation method (Hartley and Zisserman, 2000) followed by bundle-adjustment to triangulate the measurements into 3D space. In order to measure the object space error, we compute the 3D similarity transform between 3D control points and respective triangulated tie points. The alignment can be computed with a minimal number of three point correspondences, but using more than three points in a least squares manner will result in a closer alignment. Hence, we use the leave-one-out cross-validation (Kohavi, 1995) technique to assess the accuracy of our orientation results. We take seven correspondences to compute the parameters for the similarity transform and use the remaining point to estimate the object space error ϵ between observation X and ground truth point $\hat{X}$,

$$\epsilon = \sqrt{(X_x - \hat{X}_x)^2 + (X_y - \hat{X}_y)^2 + (X_z - \hat{X}_z)^2}. \quad (2)$$

Table 10 summarizes our evaluation, the error varies between 0.4 to 5.4cm, overall a RMSE of 3.2cm is achieved. Note, the reprojection error of the triangulated tie points varies between $1.1-2.5$ pixel, this is in accordance to the expected uncertainty induced by the manual tie point extraction. A subpixel accurate measurement of tie points (e.g. 0.5 pixel) would lead to a RMSE of about 1.5cm, that is close to the precision of the total station.

5. CONCLUSIONS

In this paper we demonstrated the feasibility of accurate and fast 3D scene reconstruction from unordered images captured by a UAV platform. We compared the orientation results of our fully automatic structure from motion pipeline to a standard, semi-automatic approach based on the PhotoModeler software. From our experiments we conclude that our system achieves the same accuracy in terms of reprojection error, but at a higher confidence, since many more tie points are utilized than for the semi-automatic approach. Furthermore, our method is scalable to larger datasets and allows much faster image orientation. In our experiments we achieve a speedup of about 60 over semi-automatic processing with the PhotoModeler software.

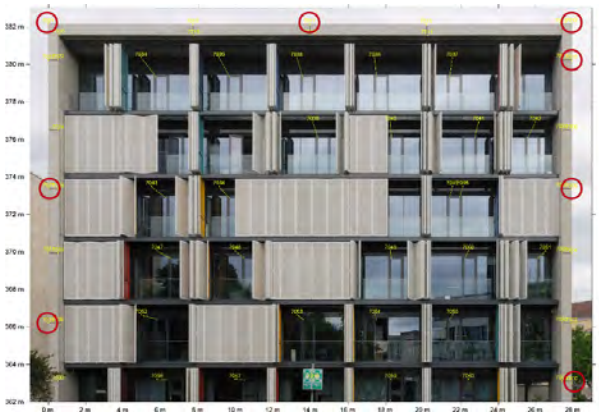


Figure 8: Orthographic projection of a building facade with the eight ground truth control points (red circles) used in our evaluation.

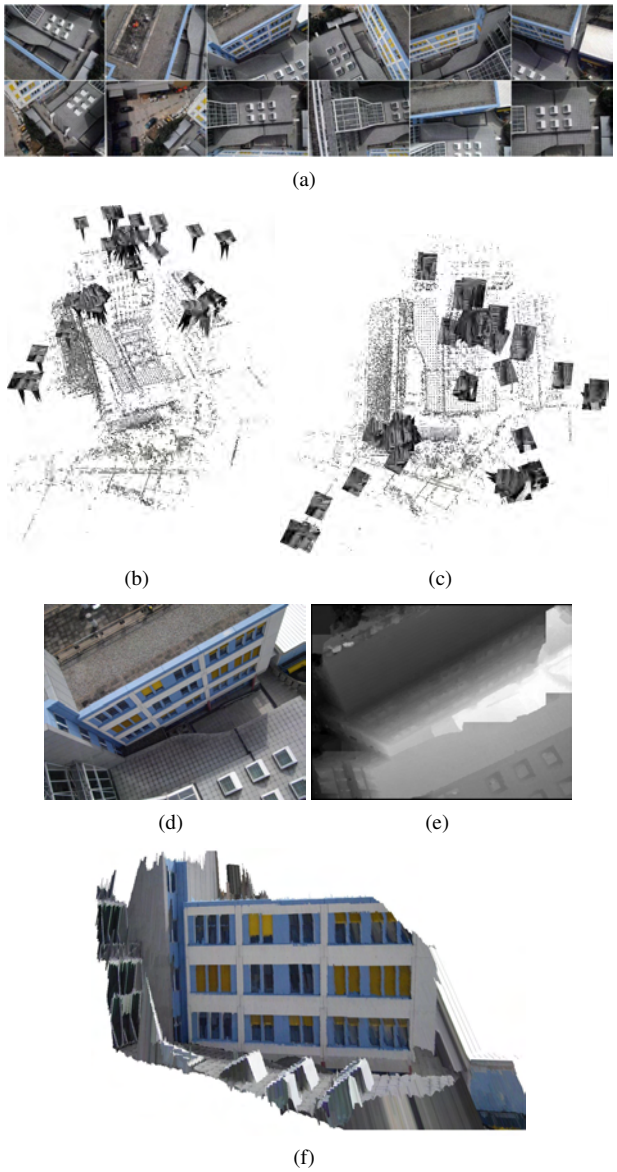


Figure 9: (a) Sample input images and (b),(c) perspective view of camera orientations and respective 3D points.(e) Input image and related depth map (f) obtained by dense matching techniques. (f) Texturized depthmap from an oblique viewpoint.

Point ID	7000	7006	7010	7012	7021	7017	7025	7029
# measurements (images)	3	6	3	3	10	3	10	6
avg. triangulation angle [°]	107.2	21.9	23.2	23.2	33.4	54.7	69.5	84.6
avg. reprojection error [pixel]	1.18	1.67	2.24	1.63	1.58	1.16	2.44	0.85
object space error [cm]	4.2	0.4	2.5	4.5	0.6	2.8	1.7	5.4

Table 10: Reprojection error and object space error determined by leave-one-out cross-validation for eight ground truth control points.

REFERENCES

Agarwal, S., Snavely, N., Simon, I., Seitz, S. M. and Szeliski, R., 2009. Building Rome in a day. In: IEEE International Conference on Computer Vision (ICCV).

Anan, C. S. and Hartley, R. I., 2008. Optimised KD-trees for fast image descriptor matching. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–8.

Chen, S. Y., Li, Y. F., Zhang, J. W. and Wang, W. L., 2008. Active Sensor Planning for Multiview Vision Tasks. Springer-Verlag.

Colomina, I., Blázquez, M., Molina, P., Parés, M. and Wis, M., 2008. Towards a new paradigm for high-resolution low-cost photogrammetry and remote sensing. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVII, Part B1, pp. 1201–1206.

Eissenbeiss, H., Nackaerts, K. and Everaerts, J., 2009. UAS for mapping & monitoring applications. In: 2009/2010 UAS Yearbook - UAS: The Global Perspective, 7th Edition, pp. 146–150.

Eugster, H. and Nebiker, S., 2009. Real-time georegistration of video streams from mini or micro UAS using digital 3D city models. In: 6th International Symposium on Mobile Mapping Technology.

Fischler, M. A. and Bolles, R. C., 1981. Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography. Communication Association and Computing Machine 24(6), pp. 381–395.

Förstner, W. and Steffen, R., 2007. Online geocoding and evaluation of large scale imagery without GPS. In: D. Fritsch (ed.), Photogrammetric Week '07, Heidelberg, pp. 243–253.

Fukunaga, K. and Narendra, P. M., 1975. A branch and bound algorithm for computing k-nearest neighbors. IEEE Transactions on Computers C-24(7), pp. 750–753.

Grenzdörffer, G., Engel, A. and Teichert, B., 2008. The photogrammetric potential of low-cost uavs in forestry and agriculture. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVII, Part B1, pp. 1207–1214.

Gruen, A. and Akca, D., 2008. Metric accuracy testing with mobile phone cameras. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVII, Part B5, pp. 729–736.

Haralick, R. M., Lee, C., Ottenberg, K. and Nölle, M., 1991. Analysis and solutions of the three point perspective pose estimation problem. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 592–598.

Hartley, R. and Zisserman, A., 2000. Multiple View Geometry in Computer Vision. Cambridge University Press.

Irschara, A., Zach, C. and Bischof, H., 2007. Towards wiki-based dense city modeling. In: Workshop on Virtual Representations and Modeling of Large-scale environments (VRML).

Irschara, A., Zach, C., Frahm, J. M. and Bischof, H., 2009. From structure-from-motion point clouds to fast location recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2599–2606.

Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: IJCAI, pp. 1137–1145.

Lowe, D., 2004. Distinctive image features from scale-invariant keypoints. Int. Journal of Computer Vision 60(2), pp. 91–110.

Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T. and Van Gool, L., 2005. A comparison of affine region detectors. Int. Journal of Computer Vision 65, pp. 43–72.

Muja, M. and Lowe, D. G., 2009. Fast approximate nearest neighbors with automatic algorithm configuration. In: A. Ranchordas and H. Araújo (eds), VISAPP (1), INSTICC Press, pp. 331–340.

Nistér, D., 2004. An efficient solution to the five-point relative pose problem. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 26(6), pp. 756–770.

Nistér, D., 2005. Preemptive RANSAC for live structure and motion estimation. Mach. Vis. App.

Nistér, D. and Stewenius, H., 2006. Scalable recognition with a vocabulary tree. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2161–2168.

Nistér, D., Naroditsky, O. and Bergen, J., 2004. Visual odometry. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 652–659.

Pollefeys, M., Koch, R. and Gool, L. V., 1999. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. Int. Journal of Computer Vision 32(1), pp. 7–25.

Pollefeys, M., Van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J. and Koch, R., 2004. Visual modeling with a hand-held camera. Int. Journal of Computer Vision 59(3), pp. 207–232.

Quaritsch, M., Stojanovski, E., Bettstetter, C., Friedrich, G., Hellwagner, H., Rinner, B., Hofbaur, M. W. and Shah, M., 2008. Collaborative microdrones: applications and research challenges. In: A. Manzalini (ed.), Autonomics, p. 38.

Scaioni, M., Barazzetti, L., Brumana, R., Cuca, B., Fassi, F. and Prandi, F., 2009. RC-heli and structure and motion techniques for the 3-D reconstruction of a milan dome spire. In: 3DARCH09.

Singhal, A., 2001. Modern information retrieval: A brief overview. IEEE Data Engineering Bulletin 24(4), pp. 35–43.

Sivic, J. and Zisserman, A., 2003. Video google: A text retrieval approach to object matching in videos. In: IEEE International Conference on Computer Vision (ICCV), pp. 1470–1477.

Snavely, N., Seitz, S. and Szeliski, R., 2006. Photo tourism: Exploring photo collections in 3D. In: Proceedings of SIGGRAPH 2006, pp. 835–846.

Steffen, R. and Förstner, W., 2008. On visual real time mapping for unmanned aerial vehicles. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVII, Part B3a, pp. 57–62.

ACKNOWLEDGEMENTS

We are grateful to Wolfgang Waagner for providing and piloting the micro-drone md4-200. This work was supported by the Austrian Science Fund (FWF) under the doctoral program Confluence of Vision and Graphics W1209 and the PEGASUS (825841) project, financed by the Austrian Research Promotion Agency (FFG).

EXTERIOR ORIENTATION OF LINE-ARRAY CCD IMAGES BASED ON QUATERNION SPHERICAL LINEAR INTERPOLATION

G. Jiang, T. Jiang *, H. Gong, X. Wang

Zhengzhou Institute of Surveying and Mapping, Information Engineering University, 66 Middle Longhai Road, Zhengzhou, Henan 450052 China - jianggw@vip.sina.com

KEY WORDS: Photogrammetry, Algorithm, CCD, Orientation, Correlation

ABSTRACT:

Since the exterior orientation elements of line-array CCD images are highly correlated, normal collinear equations that computing these elements are ill-posed and the error of the least square estimation is very large and the solution strongly depends on the initial value. For solving this problem, this paper puts forward an algorithm to compute the exterior orientation elements based on quaternion spherical linear interpolation. Firstly the quaternion is used to describe the attitude of the image, and then spherical linear interpolation is used to gain the attitude of any line in this algorithm, lastly a model of exterior orientation elements is build and is used in exterior orientation. Experimental results indicated that the method could effectively overcome the correlation problems of exterior orientation elements and the positioning accuracy is very high, and the reliability and stability of this algorithm are both independent of the initial values of exterior orientation elements.

1. INTRODUCTION

Line-array CCD images have stable geometric attributes, and it is quite meaningful to investigate the techniques of object orientation and stereo plotting by CCD images. However, line-array CCD images have a projection centre for each image line, and the orientation parameters of traditional linearization collinear equation are highly correlated. Thus, the equations to solve exterior orientation elements are greatly ill-posed. Least Square (LS) solutions, which are dependent on the initial values of exterior orientation elements, may have great errors, and the precision of orientation and mapping is greatly influenced (Wang, 1979; Qian, et.al, 1991).

Numerous investigations have been done by researchers on the precise computing the exterior orientation elements of line-array CCD images, and many methods to solve ill-posed problems are proposed (Qian, et al, 1991; Krupnik, 2000; Katiyar, et al.2003; Gupta, et al.1997), such as adding virtual error equations, combining great correlation items, iteratively solving the line and angle exterior orientation elements respectively, centralized criterion of coefficient and so on. Although the orientation precision is increased by these methods, the correlations and the ill-posed problems of normal equation are not solved. Consequently the precision is restricted. Subsequently, some scholars put forward some biased estimations (Guo, et al, 2003; Gui, et al. 2003; Wang, et al.2005), for example ridge estimation (including special and generalized ridge estimation), principal component analysis and stein estimation. However, there are various limitations in these biased estimations, and many works should be done to improve them. Moreover, for high resolution satellite CCD images (such as IKONOS images), rational function model, affine transformation method are investigated to objects positioning by many researchers (Okamoto, et al. 1999; Fraser, et al. 2002; Zhang, et al. 2004). For these methods, rigorous projection relationships between image coordinate system and ground coordinate system are not considered. This is the greatest advantage of these methods.

The ill-posed problems can be solved in two aspects. One is choosing the appropriate calculation methods. The other one is building appropriate math model. The above solutions just reduce the ill-posed problem from the calculation methods, without solving this problem in essential. In order to solve the ill-posed problem perfectly, this paper tries to build a model of exterior orientation elements using quaternion. Single image space resection based on quaternion is firstly studied and tested in Jiang's (Jiang, 2007) and Wang's (Wang, 2007) papers. Experimental results show that it can get correct solutions under a larger range of initial values than traditional way. However, frame photo is the research object in their paper, and their methods are not suit for processing the line-array CCD images. Liu's (LIU, 2008) paper extends the method to the bundle adjustment of airborne three-line images, and it behaves very well. However, his method is not suit for processing single line-array CCD images. So in this paper, when quaternion is used to describe the exterior orientation elements of the first and the last scan line of a line-array CCD image, we gain the quaternion attitude elements of any scan line through the method of quaternion spherical linear interpolation (SLERP). Then we put forward an algorithm of exterior orientation based on the quaternion SLERP (called quaternion algorithm). At last, we do some exterior orientation experiments by this algorithm.

The remainder of this paper is organized as follows: In Section 2, quaternion and quaternion spherical linear interpolation are briefly reviewed. Section 3 puts forward a SLERP model of exterior orientation elements based on quaternion, and an algorithm of exterior orientation using the above model is given. Experimental results and analysis are given in Section 4. Finally, Section 5 concludes the paper.

* Corresponding author.

2. QUATERNION SPHERICAL LINEAR INTERPOLATION

2.1 Quaternion representation of rotation

A quaternion is an ultra-complex number which expresses as $\dot{q} = q_0 + \boldsymbol{i}q_1 + \boldsymbol{j}q_2 + \boldsymbol{k}q_3$ (Zhang,1997),where q_0, q_1, q_2, q_3 are real numbers and $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ are so-called imaginary units whose products follow the rule $\boldsymbol{i}^2 = \boldsymbol{j}^2 = \boldsymbol{k}^2 = -1$, $\boldsymbol{jk} = -\boldsymbol{kj} = \boldsymbol{i}$, $\boldsymbol{ki} = -\boldsymbol{ik} = \boldsymbol{j}$, $\boldsymbol{ij} = -\boldsymbol{ji} = \boldsymbol{k}$. A quaternion also can be written as $\dot{q} = q_0 + \boldsymbol{q}$, where the number q_0 is called the real part and the sum $\boldsymbol{q} = \boldsymbol{i}q_1 + \boldsymbol{j}q_2 + \boldsymbol{k}q_3$ is called imaginary part. The vector form of quaternion is $\dot{q} = [q_0 \quad q_1 \quad q_2 \quad q_3]^{\mathrm{T}}$.The conjugate of a quaternion is $\dot{q}^* = q_0 - \boldsymbol{i}q_1 - \boldsymbol{j}q_2 - \boldsymbol{k}q_3$.The norm of a quaternion is $\|\dot{q}\| = \sqrt{\dot{q}\dot{q}^*}$.If the $\|\dot{q}\| = 1$, quaternion $\dot{q}$ is called a unit quaternion.

The detailed discussion of quaternion is in Jiang's paper. Much of that is not exited in this paper. The rotation matrix which using the quaternion can be gained from the product rule of quaternion, the formula is as follow (JIANG,2007):

$$M = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1q_2 - q_0q_3) & 2(q_1q_3 + q_0q_2) \\ 2(q_2q_1 + q_0q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2q_3 - q_0q_1) \\ 2(q_3q_1 - q_0q_2) & 2(q_2q_3 + q_0q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{bmatrix} \tag{1}$$

When the order of the rotation axis is $Y - X - Z$, we can obtain the Euler angles $(\varphi, \omega, \kappa)$ from the above rotation matrix through the following formula.

$$\left.\begin{aligned} \varphi &= \tan^{-1}\left(\frac{-M_{13}}{M_{33}}\right) \\ \omega &= \tan^{-1}(-M_{23}) \\ \kappa &= \tan^{-1}\left(\frac{M_{21}}{M_{22}}\right) \end{aligned}\right\} \tag{2}$$

2.2 Quaternion Spherical Linear Interpolation

When quaternion is used to describe the image attitude, the spherical linear interpolation of quaternion can allow us to smoothly interpolate between two image attitudes (LIU, 2008). The principle of quaternion spherical linear interpolation is illustrated in figure 1. Given two unit quaternions $\dot{q}_1$、$\dot{q}_2$ and their inclination angle θ, the unit quaternion $\dot{q}(t)$ is on the arc which connected $\dot{q}_1$ and $\dot{q}_2$, and the inclination angle of $\dot{q}(t)$ and $\dot{q}_1$ is $t\theta$. So $\dot{q}(t)$ is given by:

$$\dot{q}(t) = C_1(t)\dot{q}_1 + C_2(t)\dot{q}_2 \tag{3}$$

Where, $C_1(t), C_2(t)$ are the coefficients, and t is the interpolative variable.

$$C_1(t) = \frac{\sin(1-t)\theta}{\sin\theta} \quad C_2(t) = \frac{\sin t\theta}{\sin\theta} \tag{4}$$

$$\begin{aligned} \theta &= \arccos(\dot{q}_1 \cdot \dot{q}_2) \\ &= \arccos(q_{10}q_{20} + q_{11}q_{21} + q_{12}q_{22} + q_{13}q_{23}) \end{aligned} \tag{5}$$

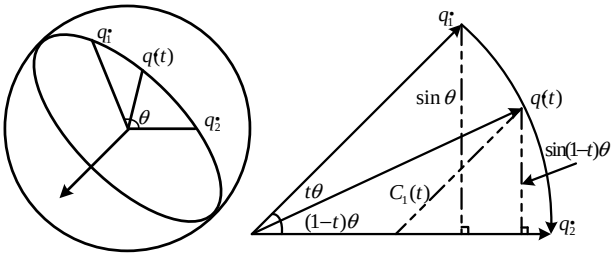


Figure 1. The outline of quaternion spherical linear interpolation

In this paper, we only give that how to obtain the expression of $C_1(t)$. From the geometry of figure 1, we can get two similarity triangles, so that.

$$\frac{C_1(t)}{\|\dot{q}_1\|} = \frac{\left\|q(\dot{t})\right\|\sin(1-t)\theta}{\|\dot{q}_1\|\sin\theta}$$

$\dot{q}_1$ and $\dot{q}(t)$ are unit quaternions, $\|\dot{q}_1\| = \left\|q(\dot{t})\right\| = 1$, so the we can obtain:

$$C_1(t) = \frac{\sin(1-t)\theta}{\sin\theta}$$

We can also obtain the expression of $C_2(t)$ similarly.

Using formula (4) in equation (3), the SLERP function of quaternion is given by

$$\dot{q}(t) = \frac{\sin(1-t)\theta}{\sin\theta}\dot{q}_1 + \frac{\sin t\theta}{\sin\theta}\dot{q}_2 \tag{6}$$

If we look upon two unit quaternions $\dot{q}_1$ and $\dot{q}_2$ as two points on the surface of a 4D sphere, quaternion SLERP will interpolate around the shortest arc that connects the two quaternions along the surface of the 4D sphere.

3. SOLUTION OF EXTERIOR ORIENTATION ELEMENTS

3.1 SLERP model of exterior orientation elements

In this paper, the projection center (i.e., X_{Si}, Y_{Si}, Z_{Si}) in the elements of exterior orientation is expressed by linear interpolation, and the unit quaternion $\dot{q}_i$ is used to describe the angular elements, the quaternion attitude of any scan line is obtained by using SLERP.

The image projection of line-array CCD image is the line central projection, so the projection center of scan line i is given by:

$$\left.\begin{aligned} X_{Si} &= X_S + \dot{X}_S y_i \\ Y_{Si} &= Y_S + \dot{Y}_S y_i \\ Z_{Si} &= Z_S + \dot{Z}_S y_i \end{aligned}\right\} \tag{7}$$

Where X_{Si}, Y_{Si}, Z_{Si} is the projection center of scan line i; y_i is the measured y coordinate of the scan line i; X_S, Y_S, Z_S is the projection center of the centre line, and $\dot{X}_S, \dot{Y}_S, \dot{Z}_S$ are the coeffifients of change in the y direction.

When quaternion is used to describe the exterior orientation elements of the first and the last scan line of a line-array CCD image, the quaternion attitude of any scan line can be obtained by using SLERP.

$$\dot{q}_i = C_1(t)\dot{q}_1 + C_2(t)\dot{q}_2 \tag{8}$$

Where $\dot{q}_i$ is the quaternion attitude of any scan line; $\dot{q}_1$ and $\dot{q}_2$ are the quaternion attitude of the first and the last scan line; $C_1(t), C_2(t)$ is given by formula (4).

Using the vector expression of $\dot{q}_i$, $\dot{q}_1$ and $\dot{q}_2$ in formula (8), $\dot{q}_i$ is given by:

$$\begin{bmatrix} q_{i0} \\ q_{i1} \\ q_{i2} \\ q_{i3} \end{bmatrix} = \begin{bmatrix} C_1(t)q_{10} + C_2(t)q_{20} \\ C_1(t)q_{11} + C_2(t)q_{21} \\ C_1(t)q_{12} + C_2(t)q_{22} \\ C_1(t)q_{13} + C_2(t)q_{23} \end{bmatrix} \tag{9}$$

The interpolative variable t is given by:

$$t = \frac{i}{n} \qquad (0 < t < 1) \tag{10}$$

Where i is the row number of the scan line; n is the total scan line number of a view image.

3.2 Linearization of collinear equation with SLERP model

In the error equation of exterior orientation based on quaternion SLERP model, the unknown elements are (X_S,Y_S,Z_S,$\dot{X}_S$,$\dot{Y}_S$, $\dot{Z}_S$,q_{10},q_{11},q_{12},q_{13},q_{20},q_{21},q_{22},q_{23}). The following is about that how to linearize the collinear equation based on quaternion SLERP model.

In the line-array CCD image, we assume that f is known, and $(x_0, y_0) = (0, 0)$. For any GCP, the coordinates are known, $\mathrm{d}X = \mathrm{d}Y = \mathrm{d}Z = 0$. The linearization error equation is given by:

$$\left.\begin{aligned} v_x =& k_{11}\mathrm{d}X_{Si} + k_{12}\mathrm{d}Y_{Si} + k_{13}\mathrm{d}Z_{Si} \\ &+ k_{14}\mathrm{d}q_{i0} + k_{15}\mathrm{d}q_{i1} + k_{16}\mathrm{d}q_{i2} + k_{17}\mathrm{d}q_{i3} - l_x \\ v_y =& k_{21}\mathrm{d}X_{Si} + k_{22}\mathrm{d}Y_{Si} + k_{23}\mathrm{d}Z_{Si} \\ &+ k_{24}\mathrm{d}q_{i0} + k_{25}\mathrm{d}q_{i1} + k_{26}\mathrm{d}q_{i2} + k_{27}\mathrm{d}q_{i3} - l_y \end{aligned}\right\} \tag{11}$$

Where

$$\begin{aligned} k_{11} &= (a_{1i}f + a_{3i}x)/\bar{Z} & k_{21} &= a_{2i}f/\bar{Z} \\ k_{12} &= (b_{1i}f + b_{3i}x)/\bar{Z} & k_{22} &= b_{2i}f/\bar{Z} \\ k_{13} &= (c_{1i}f + c_{3i}x)/\bar{Z} & k_{23} &= c_{2i}f/\bar{Z} \\ k_{14} &= 2(f + x^2/f)q_{i2} & k_{24} &= -2(q_{i1}f + q_{i3}x) \\ k_{15} &= -2(f + x^2/f)q_{i3} & k_{25} &= 2(q_{i0}f + q_{i2}x) \\ k_{16} &= -2(f + x^2/f)q_{i0} & k_{26} &= -2(q_{i3}f + q_{i1}x) \\ k_{17} &= 2(f + x^2/f)q_{i1} & k_{27} &= 2(q_{i2}f + q_{i0}x) \end{aligned}$$

$$\begin{aligned} l_x &= x - x', x' = -f(\bar{X}/\bar{Z}) \\ l_y &= 0 - y', y' = -f(\bar{Y}/\bar{Z}) \\ \bar{X} &= a_{1i}(X - X_{Si}) + b_{1i}(Y - Y_{Si}) + c_{1i}(Z - Z_{Si}) \\ \bar{Y} &= a_{2i}(X - X_{Si}) + b_{2i}(Y - Y_{Si}) + c_{2i}(Z - Z_{Si}) \\ \bar{Z} &= a_{3i}(X - X_{Si}) + b_{3i}(Y - Y_{Si}) + c_{3i}(Z - Z_{Si}) \end{aligned}$$

$$\begin{bmatrix} a_{1i} & b_{1i} & c_{1i} \\ a_{2i} & b_{2i} & c_{2i} \\ a_{3i} & b_{3i} & c_{3i} \end{bmatrix} =$$

$$\begin{bmatrix} q_{i0}^2 + q_{i1}^2 - q_{i2}^2 - q_{i3}^2 & 2(q_{i1}q_{i2} - q_{i0}q_{i3}) & 2(q_{i1}q_{i3} + q_{i0}q_{i2}) \\ 2(q_{i2}q_{i1} + q_{i0}q_{i3}) & q_{i0}^2 - q_{i1}^2 + q_{i2}^2 - q_{i3}^2 & 2(q_{i2}q_{i3} - q_{i0}q_{i1}) \\ 2(q_{i3}q_{i1} - q_{i0}q_{i2}) & 2(q_{i2}q_{i3} + q_{i0}q_{i1}) & q_{i0}^2 - q_{i1}^2 - q_{i2}^2 + q_{i3}^2 \end{bmatrix}$$

The differential equation of equation (7) is given by:

$$\left.\begin{aligned} \mathrm{d}X_{Si} &= \mathrm{d}X_S + y\mathrm{d}\dot{X}_S \\ \mathrm{d}Y_{Si} &= \mathrm{d}Y_S + y\mathrm{d}\dot{Y}_S \\ \mathrm{d}Z_{Si} &= \mathrm{d}Z_S + y\mathrm{d}\dot{Z}_S \end{aligned}\right\} \tag{12}$$

For the same reason, we can gain the differential expression of equation (9) considering that q_{i0},q_{i1},q_{i2},q_{i3} is the function of q_{10},q_{11},q_{12},q_{13} and q_{20},q_{21},q_{22},q_{23}.

$$\begin{aligned} \mathrm{d}q_{im} =& \frac{\partial q_{im}}{\partial q_{10}}\mathrm{d}q_{10} + \frac{\partial q_{im}}{\partial q_{11}}\mathrm{d}q_{11} + \frac{\partial q_{im}}{\partial q_{12}}\mathrm{d}q_{12} + \frac{\partial q_{im}}{\partial q_{13}}\mathrm{d}q_{13} \\ &+ \frac{\partial q_{im}}{\partial q_{20}}\mathrm{d}q_{20} + \frac{\partial q_{im}}{\partial q_{21}}\mathrm{d}q_{21} + \frac{\partial q_{im}}{\partial q_{22}}\mathrm{d}q_{22} + \frac{\partial q_{im}}{\partial q_{23}}\mathrm{d}q_{23} \\ &(m = 0, 1, 2, 3) \end{aligned} \tag{13}$$

In the equation (13), there are thirty-two deflection differential coefficients. Through math deduction, we can obtain the expression of these deflection differential coefficients simply. The following is the expression of the deflection differential coefficients for $\mathrm{d}q_{i0}$.

$$\begin{aligned} \frac{\partial q_{i0}}{\partial q_{10}} &= \frac{\partial \theta}{\partial q_{10}}\left(\frac{\partial C_1(t)}{\partial \theta}q_{10} + \frac{\partial C_2(t)}{\partial \theta}q_{20}\right) + C_1(t) \\ \frac{\partial q_{i0}}{\partial q_{11}} &= \frac{\partial \theta}{\partial q_{11}}\left(\frac{\partial C_1(t)}{\partial \theta}q_{10} + \frac{\partial C_2(t)}{\partial \theta}q_{20}\right) \\ \frac{\partial q_{i0}}{\partial q_{12}} &= \frac{\partial \theta}{\partial q_{12}}\left(\frac{\partial C_1(t)}{\partial \theta}q_{10} + \frac{\partial C_2(t)}{\partial \theta}q_{20}\right) \\ \frac{\partial q_{i0}}{\partial q_{13}} &= \frac{\partial \theta}{\partial q_{13}}\left(\frac{\partial C_1(t)}{\partial \theta}q_{10} + \frac{\partial C_2(t)}{\partial \theta}q_{20}\right) \\ \frac{\partial q_{i0}}{\partial q_{20}} &= \frac{\partial \theta}{\partial q_{20}}\left(\frac{\partial C_1(t)}{\partial \theta}q_{10} + \frac{\partial C_2(t)}{\partial \theta}q_{20}\right) + C_2(t) \\ \frac{\partial q_{i0}}{\partial q_{21}} &= \frac{\partial \theta}{\partial q_{21}}\left(\frac{\partial C_1(t)}{\partial \theta}q_{10} + \frac{\partial C_2(t)}{\partial \theta}q_{20}\right) \\ \frac{\partial q_{i0}}{\partial q_{22}} &= \frac{\partial \theta}{\partial q_{22}}\left(\frac{\partial C_1(t)}{\partial \theta}q_{10} + \frac{\partial C_2(t)}{\partial \theta}q_{20}\right) \\ \frac{\partial q_{i0}}{\partial q_{23}} &= \frac{\partial \theta}{\partial q_{23}}\left(\frac{\partial C_1(t)}{\partial \theta}q_{10} + \frac{\partial C_2(t)}{\partial \theta}q_{20}\right) \end{aligned} \tag{14}$$

Where

$$\frac{\partial C_1(t)}{\partial \theta} = \frac{(1-t)\cos(1-t)\theta\sin\theta - \sin(1-t)\theta\cos\theta}{\sin^2\theta}$$

$$\frac{\partial C_2(t)}{\partial \theta} = \frac{t\cos t\theta\sin\theta - \sin t\theta\cos\theta}{\sin^2\theta}$$

$$\begin{aligned} \frac{\partial \theta}{\partial q_{10}} &= -\frac{q_{20}}{\sqrt{1-(\dot{q}_1\cdot\dot{q}_2)^2}} & \frac{\partial \theta}{\partial q_{20}} &= -\frac{q_{10}}{\sqrt{1-(\dot{q}_1\cdot\dot{q}_2)^2}} \\ \frac{\partial \theta}{\partial q_{11}} &= -\frac{q_{21}}{\sqrt{1-(\dot{q}_1\cdot\dot{q}_2)^2}} & \frac{\partial \theta}{\partial q_{21}} &= -\frac{q_{11}}{\sqrt{1-(\dot{q}_1\cdot\dot{q}_2)^2}} \\ \frac{\partial \theta}{\partial q_{12}} &= -\frac{q_{22}}{\sqrt{1-(\dot{q}_1\cdot\dot{q}_2)^2}} & \frac{\partial \theta}{\partial q_{22}} &= -\frac{q_{12}}{\sqrt{1-(\dot{q}_1\cdot\dot{q}_2)^2}} \\ \frac{\partial \theta}{\partial q_{13}} &= -\frac{q_{23}}{\sqrt{1-(\dot{q}_1\cdot\dot{q}_2)^2}} & \frac{\partial \theta}{\partial q_{23}} &= -\frac{q_{13}}{\sqrt{1-(\dot{q}_1\cdot\dot{q}_2)^2}} \end{aligned}$$

Using the deflection differential coefficients of equation (13) in error equation (11), the error equation of exterior orientation based on quaternion SLERP model is given by:

$$\left.\begin{aligned} v_x =& k'_{11}\mathrm{d}X_S + k'_{12}\mathrm{d}Y_S + k'_{13}\mathrm{d}Z_S \\ &+ k'_{14}\mathrm{d}q_{10} + k'_{15}\mathrm{d}q_{11} + k'_{16}\mathrm{d}q_{12} + k'_{17}\mathrm{d}q_{13} \\ &+ k'_{18}\mathrm{d}q_{20} + k'_{19}\mathrm{d}q_{21} + k'_{110}\mathrm{d}q_{22} + k'_{111}\mathrm{d}q_{23} \\ &+ yk'_{11}\mathrm{d}\dot{X}_S + yk'_{12}\mathrm{d}\dot{Y}_S + yk'_{13}\mathrm{d}\dot{Z}_S - l_x \\ v_x =& k'_{21}\mathrm{d}X_S + k'_{22}\mathrm{d}Y_S + k'_{23}\mathrm{d}Z_S \\ &+ k'_{24}\mathrm{d}q_{10} + k'_{25}\mathrm{d}q_{11} + k'_{26}\mathrm{d}q_{12} + k'_{27}\mathrm{d}q_{13} \\ &+ k'_{28}\mathrm{d}q_{20} + k'_{29}\mathrm{d}q_{21} + k'_{210}\mathrm{d}q_{22} + k'_{211}\mathrm{d}q_{23} \\ &+ yk'_{21}\mathrm{d}\dot{X}_S + yk'_{22}\mathrm{d}\dot{Y}_S + yk'_{23}\mathrm{d}\dot{Z}_S - l_y \end{aligned}\right\} \tag{15}$$

Where

$$\begin{aligned} k'_{11} &= k_{11} = (a_{1i}f + a_{3i}x)/\bar{Z} & k'_{21} &= k_{21} = a_{2i}f/\bar{Z} \\ k'_{12} &= k_{12} = (b_{1i}f + b_{3i}x)/\bar{Z} & k'_{22} &= k_{22} = b_{2i}f/\bar{Z} \\ k'_{13} &= k_{13} = (c_{1i}f + c_{3i}x)/\bar{Z} & k'_{23} &= k_{23} = c_{2i}f/\bar{Z} \end{aligned}$$

$$
\begin{aligned}
k'_{14} &= k_{14}\frac{\partial q_{i0}}{\partial q_{10}} + k_{15}\frac{\partial q_{i1}}{\partial q_{10}} + k_{16}\frac{\partial q_{i2}}{\partial q_{10}} + k_{17}\frac{\partial q_{i3}}{\partial q_{10}}\\
k'_{15} &= k_{14}\frac{\partial q_{i0}}{\partial q_{11}} + k_{15}\frac{\partial q_{i1}}{\partial q_{11}} + k_{16}\frac{\partial q_{i2}}{\partial q_{11}} + k_{17}\frac{\partial q_{i3}}{\partial q_{11}}\\
k'_{16} &= k_{14}\frac{\partial q_{i0}}{\partial q_{12}} + k_{15}\frac{\partial q_{i1}}{\partial q_{12}} + k_{16}\frac{\partial q_{i2}}{\partial q_{12}} + k_{17}\frac{\partial q_{i3}}{\partial q_{12}}\\
k'_{17} &= k_{14}\frac{\partial q_{i0}}{\partial q_{13}} + k_{15}\frac{\partial q_{i1}}{\partial q_{13}} + k_{16}\frac{\partial q_{i2}}{\partial q_{13}} + k_{17}\frac{\partial q_{i3}}{\partial q_{13}}\\
k'_{18} &= k_{14}\frac{\partial q_{i0}}{\partial q_{20}} + k_{15}\frac{\partial q_{i1}}{\partial q_{20}} + k_{16}\frac{\partial q_{i2}}{\partial q_{20}} + k_{17}\frac{\partial q_{i3}}{\partial q_{20}}\\
k'_{19} &= k_{14}\frac{\partial q_{i0}}{\partial q_{21}} + k_{15}\frac{\partial q_{i1}}{\partial q_{21}} + k_{16}\frac{\partial q_{i2}}{\partial q_{21}} + k_{17}\frac{\partial q_{i3}}{\partial q_{21}}\\
k'_{110} &= k_{14}\frac{\partial q_{i0}}{\partial q_{22}} + k_{15}\frac{\partial q_{i1}}{\partial q_{22}} + k_{16}\frac{\partial q_{i2}}{\partial q_{22}} + k_{17}\frac{\partial q_{i3}}{\partial q_{22}}\\
k'_{111} &= k_{14}\frac{\partial q_{i0}}{\partial q_{23}} + k_{15}\frac{\partial q_{i1}}{\partial q_{23}} + k_{16}\frac{\partial q_{i2}}{\partial q_{23}} + k_{17}\frac{\partial q_{i3}}{\partial q_{23}}\\
k'_{24} &= k_{24}\frac{\partial q_{i0}}{\partial q_{10}} + k_{25}\frac{\partial q_{i1}}{\partial q_{10}} + k_{26}\frac{\partial q_{i2}}{\partial q_{10}} + k_{27}\frac{\partial q_{i3}}{\partial q_{10}}\\
k'_{25} &= k_{24}\frac{\partial q_{i0}}{\partial q_{11}} + k_{25}\frac{\partial q_{i1}}{\partial q_{11}} + k_{26}\frac{\partial q_{i2}}{\partial q_{11}} + k_{27}\frac{\partial q_{i3}}{\partial q_{11}}\\
k'_{26} &= k_{24}\frac{\partial q_{i0}}{\partial q_{12}} + k_{25}\frac{\partial q_{i1}}{\partial q_{12}} + k_{26}\frac{\partial q_{i2}}{\partial q_{12}} + k_{27}\frac{\partial q_{i3}}{\partial q_{12}}\\
k'_{27} &= k_{24}\frac{\partial q_{i0}}{\partial q_{13}} + k_{25}\frac{\partial q_{i1}}{\partial q_{13}} + k_{26}\frac{\partial q_{i2}}{\partial q_{13}} + k_{27}\frac{\partial q_{i3}}{\partial q_{13}}\\
k'_{28} &= k_{24}\frac{\partial q_{i0}}{\partial q_{20}} + k_{25}\frac{\partial q_{i1}}{\partial q_{20}} + k_{26}\frac{\partial q_{i2}}{\partial q_{20}} + k_{27}\frac{\partial q_{i3}}{\partial q_{20}}\\
k'_{29} &= k_{24}\frac{\partial q_{i0}}{\partial q_{21}} + k_{25}\frac{\partial q_{i1}}{\partial q_{21}} + k_{26}\frac{\partial q_{i2}}{\partial q_{21}} + k_{27}\frac{\partial q_{i3}}{\partial q_{21}}\\
k'_{210} &= k_{24}\frac{\partial q_{i0}}{\partial q_{22}} + k_{25}\frac{\partial q_{i1}}{\partial q_{22}} + k_{26}\frac{\partial q_{i2}}{\partial q_{22}} + k_{27}\frac{\partial q_{i3}}{\partial q_{22}}\\
k'_{211} &= k_{24}\frac{\partial q_{i0}}{\partial q_{23}} + k_{25}\frac{\partial q_{i1}}{\partial q_{23}} + k_{26}\frac{\partial q_{i2}}{\partial q_{23}} + k_{27}\frac{\partial q_{i3}}{\partial q_{23}}
\end{aligned}
$$

The expression of l_x and l_y is as the same as that in equation (11).

In the result of the above linearization equation, the form of the coefficients is fit to program. Comparing with the traditional Euler angle model, it successfully avoids a great deal of computation of trigonometric functions.

3.3 Computation of exterior orientation elements

The error equation (15) can be written in the matrix form.

$$\boldsymbol{V} = \boldsymbol{C}\boldsymbol{\delta}_x + \boldsymbol{l} \tag{16}$$

Where

$$
\begin{aligned}
&\boldsymbol{V} = [v_x\ v_y]^{\mathrm{T}}\\
&\boldsymbol{C} = \begin{bmatrix} k'_{11} & k'_{12} & \cdots & k'_{111} & yk'_{11} & yk'_{12} & yk'_{13} \\ k'_{21} & k'_{22} & \cdots & k'_{211} & yk'_{21} & yk'_{22} & yk'_{23} \end{bmatrix}\\
&\boldsymbol{\delta}_x = [\mathrm{d}X_S\ \mathrm{d}Y_S\ \mathrm{d}Z_S\ \mathrm{d}q_{10}\ \mathrm{d}q_{11}\ \mathrm{d}q_{12}\ \mathrm{d}q_{13}\\
&\qquad \mathrm{d}q_{20}\ \mathrm{d}q_{21}\ \mathrm{d}q_{22}\ \mathrm{d}q_{23}\ \mathrm{d}\dot{X}_S\ \mathrm{d}\dot{Y}_S\ \mathrm{d}\dot{Z}_S]^{\mathrm{T}}\\
&\boldsymbol{l} = [l_x\ l_y]^{\mathrm{T}}
\end{aligned}
$$

Because of the unit quaternion which is used to describe the attitude, there are two conditions in the error equations.

$$
\left.\begin{aligned}
q_{10}^2 + q_{11}^2 + q_{12}^2 + q_{13}^2 = 1\\
q_{20}^2 + q_{21}^2 + q_{22}^2 + q_{23}^2 = 1
\end{aligned}\right\}
$$

The two equations also should be linearized, and the linearization equation is given by:

$$
\begin{aligned}
&q_{10}\mathrm{d}q_{10} + q_{11}\mathrm{d}q_{11} + q_{12}\mathrm{d}q_{12} + q_{13}\mathrm{d}q_{13} + w_1 = 0\\
&w_1 = (q_{10}^2 + q_{11}^2 + q_{12}^2 + q_{13}^2 - 1)\\
&q_{20}\mathrm{d}q_{20} + q_{21}\mathrm{d}q_{21} + q_{22}\mathrm{d}q_{22} + q_{23}\mathrm{d}q_{23} + w_2 = 0\\
&w_2 = (q_{20}^2 + q_{21}^2 + q_{22}^2 + q_{23}^2 - 1)
\end{aligned}
$$

The above equations can be written as follow:

$$\boldsymbol{B}_x\boldsymbol{\delta}_x + \boldsymbol{W} = 0 \tag{17}$$

Where

$$
\begin{aligned}
&\boldsymbol{B}_x = \begin{bmatrix} 0\ 0\ 0\ q_{10}\ q_{11}\ q_{12}\ q_{13}\ 0\ 0\ 0\ 0\ 0\ 0\ 0 \\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ q_{20}\ q_{21}\ q_{22}\ q_{23}\ 0\ 0\ 0 \end{bmatrix}\\
&\boldsymbol{w} = [w_1\ w_2]^{\mathrm{T}}
\end{aligned}
$$

When there are n points, the total error equations is:

$$
\left.\begin{aligned}
\boldsymbol{V} = \boldsymbol{C}\boldsymbol{\delta}_x + \boldsymbol{l}\\
\boldsymbol{B}_x\boldsymbol{\delta}_x + \boldsymbol{W} = 0
\end{aligned}\right\} \tag{18}
$$

Where $\boldsymbol{V}$ and $\boldsymbol{l}$ are the matrix with $2n \times 1$; $\boldsymbol{C}$ is the matrix with $2n \times 10$; n is the number of ground control points (GCPs).

Supposed the weight matrix of observed value is $\boldsymbol{P}$, we can gain the solution of equation (18) according parameter adjustment with constraints.

$$\boldsymbol{Y} = -\boldsymbol{N}^{-1}\boldsymbol{W}_Y \tag{19}$$

Where

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{\delta}_x \\ \boldsymbol{K} \end{bmatrix}, \boldsymbol{N} = \begin{bmatrix} \boldsymbol{C}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{C} & \boldsymbol{B}_x^{\mathrm{T}} \\ \boldsymbol{B}_x & 0 \end{bmatrix}, \boldsymbol{W}_Y = \begin{bmatrix} \boldsymbol{C}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{C} \\ \boldsymbol{W} \end{bmatrix} \tag{20}$$

When the number of GCPs $n \geq 6$, given the initial value of line and attitude elements $\boldsymbol{\delta}_x^0$, we can obtain the most probable value of the line and attitude elements $\boldsymbol{\delta}_x$ iteratively through equation (19) until that the result is smaller than the tolerance.

3.4 Summary of the algorithm

The complete algorithm works as follows:

(1) Input the base data, including observed value of image points and the ground coordinate of the corresponding GCPs.

(2) Determine the initial values of exterior orientation elements $\boldsymbol{\delta}_x^0$. In this quaternion algorithm, the initial values has not special requirement. We just can give the initial values as $X_{S0} = \bar{X}_{GCP}$, $Y_{S0} = \bar{Y}_{GCP}$, $Z_{S0} = H$, $q_{10} = 1$, $q_{11} = q_{12} = q_{13} = 0$, $q_{20} = 1$, $q_{21} = q_{22} = q_{23} = 0$, $\dot{X}_S = \dot{Y}_S = \dot{Z}_S = 0$,where $\bar{X}_{GCP}$,$\bar{Y}_{GCP}$ are the average value of GCPs ground coordinate. Then begin the iteration.

(3) Compute exterior orientation elements of every GCP using equation (7) and (9), and calculate the quaternion rotation matrix using equation (1).

(4) Compute the matrix$\boldsymbol{C}$,$\boldsymbol{l}$ and weight matrix $\boldsymbol{P}$ using error equation (16), and estimate value of the least squares prediction $\boldsymbol{\delta}_x$.

(5) Update the exterior orientation elements with formula $\boldsymbol{\delta}_x^1 = \boldsymbol{\delta}_x^0 + \boldsymbol{\delta}_x$, and check that whether the result $\boldsymbol{\delta}_x$ is less than the tolerance or not. If it is less than the tolerance, the iteration ends, and the result $\boldsymbol{\delta}_x^1$ is the estimation value of the exterior orientation elements. Otherwise, repeat step 3 to step 4 until the result of exterior orientation elements is less than the tolerance.

4. EXPERIMENTS AND RESULTS

The exterior orientation process of line-array CCD images is to obtain the exterior orientation elements of the images using GCPs and their image points. In order to verify the efficiency of this quaternion method, experiments are done on the solution of exterior orientation elements by various methods.

The experiment data used in this paper are two SPOT images (Level 1A) of TangShan area in china: Image_01 and Image_02, which constitute a single stereo. The size of the pixel is 13μm×13μm, the focal length of the camera is 1082 mm, and the size of the image is 6000×6000. The flight height is about 830 000.0 m, the swath width is 60km×60km, and the overlap of these two images exceeds 80%. The ground coordinate uses Gaussian coordinate system, and the longitude of central meridian is 117 degree. The ground altitude difference is 0～500 m.

Figure 2 is the outline of GCPs which is distributed evenly in images. The control points are measured from 1:50,000 topographic maps. For Image_01 and Image_02, 19 control points are selected respectively, with 13 orientation points and 6 checking points. There are 13 homologous points in the two images.

We use the following methods in our experiments.

Method 1: Our algorithm

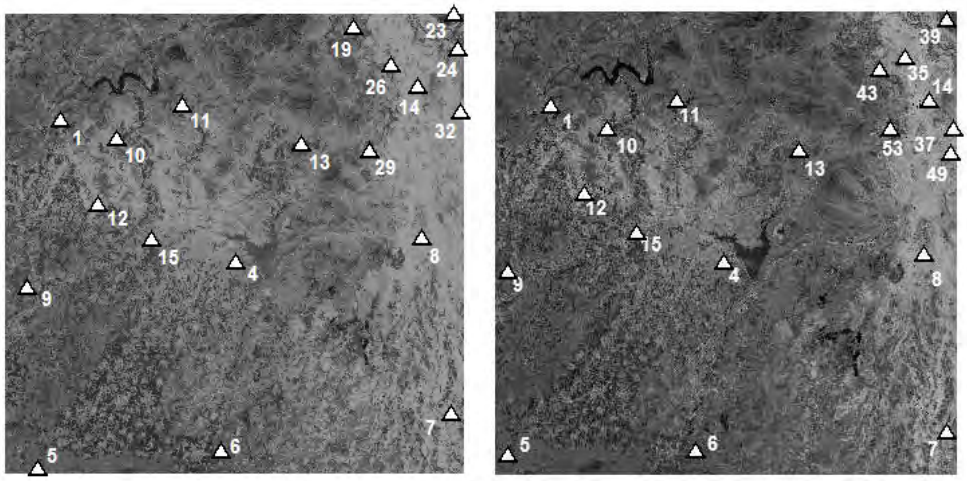
Method 2: LS estimation (Qian, 1991).

Method 3: General ridge estimation (Guo, 2003).

In order to validate the independence of the initial value of our method, we use the coordinate of GCPs to compute the initial value in our experiment. So, we consider that there is not any priori information of images, and the initial values are as follow: $X_{S0}=\bar{X}_{GCP}$, $Y_{S0}=\bar{Y}_{GCP}$, $Z_{S0}=H$, $q_{10}=1$, $q_{11}=q_{12}=q_{13}=0$, $q_{20}=1$, $q_{21}=q_{22}=q_{23}=0$, $\dot{X}_S=\dot{Y}_S=\dot{Z}_S=0$.

In our experiments, we use the analytical method to analyse the precision. The solution precision of exterior orientation elements can be evaluated by the following two checking methods. One is computing the reprojection error. The other is computing ground points coordinate by space intersection, and then calculating the mean square error.

In table 3 and table 4, $\boldsymbol{v}_{ox}$,$\boldsymbol{v}_{oy}$,$\boldsymbol{v}_{oX}$,$\boldsymbol{v}_{oY}$,$\boldsymbol{v}_{oZ}$are the precision (mean square error) for the image coordinates x,y and the ground coordinates X,Y,Z of orientation points respectively; while $\boldsymbol{v}_{ex}$,$\boldsymbol{v}_{cy}$,$\boldsymbol{v}_{eX}$,$\boldsymbol{v}_{eY}$,$\boldsymbol{v}_{eZ}$ are the precision (mean square error) for the image coordinates x,y and the ground coordinates X,Y,Z of the checking points respectively.



(a) (b)

Figure 2. The outline of GCPs : (a) Image_01 (b) Image_02

Method	Mean Square error for the coordinates of orientation points					Mean Square error for the coordinates of checking points				
	$\boldsymbol{v}_{ox}$/Pixel	$\boldsymbol{v}_{oy}$/Pixel	$\boldsymbol{v}_{oX}$/m	$\boldsymbol{v}_{oY}$/m	$\boldsymbol{v}_{oZ}$/m	$\boldsymbol{v}_{ex}$/Pixel	$\boldsymbol{v}_{cy}$/Pixel	$\boldsymbol{v}_{eX}$/m	$\boldsymbol{v}_{eY}$/m	$\boldsymbol{v}_{eZ}$/m
Method 1	0.91	0.87	7.68	10.16	17.08	1.35	1.22	13.40	10.13	21.38
Method 2	1.68	1.58	20.03	22.87	84.92	1.90	1.82	25.87	31.89	96.32
Method 3	0.91	0.93	9.23	14.92	41.94	1.61	1.12	20.79	13.43	63.35

Table 3. Orientation precision of Image_01 under various methods

Method	Mean Square error for the coordinates of orientation points					Mean Square error for the coordinates of checking points				
	$\boldsymbol{v}_{ox}$/Pixel	$\boldsymbol{v}_{oy}$/Pixel	$\boldsymbol{v}_{oX}$/m	$\boldsymbol{v}_{oY}$/m	$\boldsymbol{v}_{oZ}$/m	$\boldsymbol{v}_{ex}$/Pixel	$\boldsymbol{v}_{cy}$/Pixel	$\boldsymbol{v}_{eX}$/m	$\boldsymbol{v}_{eY}$/m	$\boldsymbol{v}_{eZ}$/m
Method 1	0.89	0.91	7.68	10.16	17.08	1.50	1.27	13.40	10.13	21.38
Method 2	1.52	1.58	20.03	22.87	84.92	1.71	1.49	25.87	31.89	96.32
Method 3	1.08	1.59	9.23	14.92	41.94	1.25	2.14	20.79	13.43	63.35

Table 4. Orientation precision of Image_02 under various methods

From table 3 and table 4, it can be seen that the quaternion algorithm can obtain exterior orientation elements successfully. The plane positioning precision is about 1 GSD, and better than that of the LS estimation and the general ridge estimation. However, the height positioning precision is less admirable, we think that the main reason is the influence of the measure error for GCPs. Of course, due to the initial values which gained by GCPs only, the positioning precision of the LS estimation and the general ridge estimation is not very high, and that show our algorithm is independence on the initial values. At the same time, special calculation measures are not required in the computing process, which shows that the influence of ill-posed problem is weaken effectively by this quaternion algorithm. However, because the attitude quaternions which used to describe the exterior orientation elements of the first and the last scan line of a line-array CCD image are unknown numbers, the redundant parameters is also exited, and deep researches are needed.

5. CONCLUSIONS

Line-array CCD Images are widely used in target positioning of remote sensing imagery. However, the orientation parameters of traditional linearization collinear equation are highly correlated because of such factors as big flying height, small angle of coverage, narrow light beam and so on. Thus, it is fairly difficult to gain steady and precise solutions. We use quaternion to describe the attitude elements in this paper, and then build a model of exterior orientation elements. It can obtain reliable and precise positioning results. Compared with traditional solution methods, it successfully avoids a great deal of computation of trigonometric functions and the potential unstable factors, and hasn't special requirement for initial values. So it is quite suit for exterior orientation of line-array CCD Images without good initial values. At present, there are many space photographs without ephemeris data, and how to use these images to position is a challengeable work in photogrammetry. The quaternion algorithm in this paper provides a method for finding the exterior orientation elements of these line-array CCD images, and provides precise initial values for consequent applications.

REFERENCE

Amnon krupnik. 2000. Accuracy assessment of automatically derived digital elevation models from spot image. *Photogrammetric Engineering and Remote Sensing*. 66(8), pp.1017-1023.

Fraser C S, Hanley H B, Yamakawa T. 2002, Three-Dimensional Geopositioning Accuracy of Ikonos Imagery. *Photogrammetric Record*, 17(99), pp.465-479.

Fuzhen Zhang. 1997. Quaternions and Matrices of Quaternions .*Linear Algebra and its Applications.* 252, pp.21-57.

Gui Qingming, Guo Haitao, Guo Jianfeng, et al. 2003. Computing the Exterior Orientation Elements of One-line Scanner Satellitic Imagery By Using Principal Component Estimation, *Remote Sensing Technology and Application*, 18(1), pp.14-18. (In Chinese.)

Guo Haitao, Zhang Baoming, Gui Qingming. 2003. Application of generalized ridge estimate to computing the exterior orientation elements of satellite linear array scanner imagery, *Geomatics and Information Science of Wuhan University*, 28(4). pp.444-447. (In Chinese.)

Jiang Gang-wu, Jiang Ting, Wang Yong, et al. 2007,5. Space Resection Indepdent of Initial Value Based on Unit Quaternions. *Acta Geodaetica et Cartographica Sinica*, 36(2), pp.169-175. (In Chinese.)

Katiyar S.K, Onkar Dikshit, Krishna Kumar. 2003. Linear pushbroom model for IRS-1C/D satellite imaging geometry. *IEEE International of Geoscience and Remote Sensing Symposium*, vol 6, pp.3613-3615.

Liu Jun, Wang Donhong, Zhang Yongsheng. 2008. Triangulation of Airborne Three-Line Images using Unit Quaternion. *The International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences.* Beijing, China Vol. XXXVII. Part B1. pp.573-578.

Okamoto A, Ono T, et al. 1999. Geometric Characteristics of Alternative Triangulation Models for Satellite Imagery. *ASPRS 1999 Aunual Conference Proceedings*. Oregon., pp. 64-72.

Qian Zhengbo, Liu Jingyu, Xiao Guochao. 1991. *Space Photogrammetry*. Publishing House of Surveying and Mapping. Beijing. (In Chinese.)

Rajiv Gupta, Richard I. Hartley, 1997. Linear pushbroom cameras, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 963-975.

Wang Tao, Zhang Yan, Xu Qing, et al. 2005, A New Method for Linear Pushbroom Imagery Exterior Orientation. *Acta Geodaetica et Cartographica Sinica*, 34(1), pp.35-39. (In Chinese.)

Wang Yong, Jiang Ting, Jiang Gangwu, et al. 2007, Space Resection of Single Image Based on the Description of Unit Quaternions . *Journal of Zhenzhou Institute of Surveying and Mapping*. 24(2), pp.133-135. (In Chinese.)

Wang Zhizhuo. 1979. *The Theory of Photogrammetry*, Publishing House of Surveying and Mappin. Beijing.

Zhang Yongsheng, Gong Danchao. 2004. *Application of high resolution remote sensing satellites*, Publishing House of Science. Beijing,. (In Chinese.)

ACKNOWLEDGEMENTS

This work is supported by the National Natural Science Foundation of China (No. 40901246).

THEORETICAL FRAMEWORKS OF REMOTE SENSING SYSTEMS BASED ON COMPRESSIVE SENSING

L. Jiying *, Z. Jubo, Y. Fengxia, Z. Zenghui

Dept. of Mathematics and Systems Science, National Univ. of Defense Technology, Changsha 410073, China - gene0572@163.com

KEY WORDS: Sampling, Mathematics, Optical, SAR, Representation

ABSTRACT:

As an application of Compressive Sensing (CS) in remote sensing area, the theoretical frameworks of SAR and optical imaging system based on CS are investigated. The processes of data acquisition are mathematically described. After that the sparse representation of images corresponding to the two systems are also presented. Finally, the spare recovery is employed to retrieve images. Numerical simulations validated the feasibility of such imaging systems.

1. INTRODUCTION

Compressive Sensing (CS) provides us with a new theory for signal/image acquisition. Employing this theory, we can reconstruct signals with equivalent or better qualities (e.g.: resolution, SNR, etc.) by using less sensors, slower sampling rate, smaller data size or lower power consumption than that required in traditional system. Instead of uniform and periodical samples, CS measurements are formed by the inner products of signals with certain sensing matrix. The sparsity of signal is exploited to accurate recovery, and the measurements utilized are no longer depending on the signal's bandwidth but on the signal's sparsity. Generally speaking, the dimension of measurement vector is logarithmically with the dimension of signal (Candes, 2006; Candes, 2006b; Candes, 2006c; Donoho, 2006).

In this paper, we focus our research on the application of CS in remote sensing systems, and especially for optical and synthetic aperture radar (SAR) imaging system. In the area of optical compressive imaging, Baraniuk's group realized a Rice Single-Pixel CS Camera (Duarte, 2008), in which the Digital Micro-mirror Device (DMD) is served as sensing matrix. This camera suffers from an inherent inefficiency: sequential measurements are needed. But in many scenarios, there is a high-speed movement between the imaging sensor and target (such as spaceborne remote sensing), and the sequential multiple measurements is not permitted. A. Stern and B. Javidi proposed a random projection imaging system (Stern, 2007), in which the measurements are obtained within a single exposure by using a random phase mask. Enjoying sparse recovery, the more object pixels may be reconstructed and visualized than the number of pixels of the image sensor. But the design of sensing matrix and sparse recovery algorithm desire much improvement. Meanwhile, in the area of radar imaging, Baraniuk introduced a compressive radar imaging system (Baraniuk, 2007), but the simulation is too simple and far away from practical application. Besides, J. Romberg proposed a sampling strategy (Romberg, 2008) based on "random convolution", and discussed its application in radar and Fourier optics conceptually.

We investigated the theoretical frameworks of compressive radar and optical imaging systems, which involves: 1) mathematically reformatted the processes of data acquisition of SAR and optical imaging in the form of linear system, and then the imaging process becomes the inverse problem. Furthermore, a process called random phase modulation is specially designed for CS optical imaging system. 2) The sensing and sparse representation matrices are chosen according to the characteristics of data acquisition and images from SAR or optical imaging systems respectively. Due to the large data scale of two-dimensional imaging problem, we also give attentions to the computational efficiency of sparse recovery. Finally, in each system, numerical simulations are conducted to validate the feasibility. Especially, the data for CS SAR imaging is generated from professional electromagnetic scattering computing software which is similar to real SAR raw data.

2. COMPRESSIVE SAMPLING AND SPARSE RECOVERY

Different from the traditional uniform and periodical samples, the measurements in CS are the projections of the signal onto the sensing matrix, i.e.

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x}_0 + \boldsymbol{\varepsilon} \tag{1}$$

where $\mathbf{x}_0$ is the N-dimensional signal, $\mathbf{y}$ is the M-dimensional measurement, $\mathbf{\Phi}$ is the sensing matrix, $\boldsymbol{\varepsilon}$ stands for the noise in the data acquisition process, and its variance is σ^2. The dimension of $\mathbf{y}$ is far smaller than $\mathbf{x}_0$, i.e.: $M \ll N$. In order to reconstruct original signal, the sparsity of $\mathbf{x}_0$ is required, that is, with a representation matrix $\mathbf{\Psi}$, we can decompose $\mathbf{x}_0$ as $\mathbf{x}_0 = \mathbf{\Psi}\boldsymbol{\alpha}_0$, where the coefficient $\boldsymbol{\alpha}_0$ has at most K non-zero (or almost non-zero) components. Substitutes the signal's sparse representation into Eq. (1), then

$$\mathbf{y} = \mathbf{\Phi}\mathbf{\Psi}\boldsymbol{\alpha}_0 + \boldsymbol{\varepsilon} \triangleq \mathbf{\Theta}\boldsymbol{\alpha}_0 + \boldsymbol{\varepsilon}, \tag{2}$$

where $\mathbf{\Theta}$ is a matrix compound by the representation and sensing matrix.

* Corresponding author.

The coefficient $\boldsymbol{\alpha}_0$, so the signal $\mathbf{x}_0$, can be recovered by solving a convex program

$$\min \|\boldsymbol{\alpha}\|_1 \quad \text{s.t.} \quad \|\mathbf{y}-\boldsymbol{\Theta}\boldsymbol{\alpha}\|_2^2 \le \sigma^2, \tag{3}$$

given the matrix $\boldsymbol{\Theta}$ obyes a Restricted Isometry Property (RIP) (Candes, 2006c).

Roughly speaking, the RIP has restrained the theoretical lower bound of number of measurements. For $M \times N$ -dimensional random sensing matrices whose entries are independently generated according to the Gaussian or Bernoulli distribution, when the sparsity of signal K and the dimension of signal N are given, the number of measurements M must obey $M \ge K \log N$.

Many algorithms have been developed to handle the optimization in Eq. (3). Basis Pursuit (BP) (Chen, 1999) is one of the first proposed methods. This method enjoys rigorous proofs of exact reconstruction, but suffers from heavy computationally burdens and can not be used in two-dimensional imaging which involves data with large scale. Algorithms proposed recently have improved computationally complexity without loss of precision. These algorithms include: Gradient Projections for Sparse Reconstruction (Figueiredo, 2007), Sparse Reconstruction by separable approximation (Wright, 2009), Spectral Projected Gradient (Van Den Berg, 2008), Fixed Point Continuation method (Hale, 2008) and its modification (Wen, 2008), Bregman iteration (Cai, 2008; Osher, 2008) etc.

The sparse recovery algorithm employed in this paper is Fixed Point Continuation (FPC) algorithm, which can solve large scale problem. We briefly describe it in Algorithm 1.

Algorithm 1 (FPC algorithm)

Input: $\mathbf{y}, \boldsymbol{\Theta}$; Parameter: $\bar{\mu}, \tau, \eta, xtol$

Initializtion: $\boldsymbol{\alpha}_k = 0, \boldsymbol{\alpha} = \tau\boldsymbol{\Theta}^T\mathbf{y}, \mu = \dfrac{1}{\|\boldsymbol{\Theta}^T\mathbf{y}\|_\infty}$

while $\mu \le \bar{\mu}$

  while $\dfrac{\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_k\|_2}{\|\boldsymbol{\alpha}_k\|_2} > xtol$

    $\boldsymbol{\alpha}_k = \boldsymbol{\alpha}$

    $\mathbf{g} = \boldsymbol{\Theta}^T(\boldsymbol{\Theta}\mathbf{x}-\mathbf{y})$

    $\mathbf{b} = \boldsymbol{\alpha} - \tau\mathbf{g}$

    $\boldsymbol{\alpha} = \operatorname{sgn}(\mathbf{b}) \cdot \max\left\{|\mathbf{b}| - \dfrac{\tau}{\mu}, 0\right\}$

  end while

  $\mu = \min\{\eta\mu, \bar{\mu}\}$

end while

Output: $\boldsymbol{\alpha}, \mathbf{x}_0 = \boldsymbol{\Psi}\boldsymbol{\alpha}$

The value of relevant parameter recommend are: $\bar{\mu} = 1\times10^{-6}$, $\tau = 1.999$, $\eta = 4$ and $xtol = 1\times10^{-5}$.

3. RADAR IMAGING BASED ON CS

Radar image is a reflection of the electromagnetic scattering of the illuminated target. The process of "Radar Imaging" is to obtain the scattering coefficients from radar echo (i.e. the inverse process of radar echo generation). Supposing $\sigma(p_x, p_y)$ a two-dimensional function which describes the scattering coefficients, and then the radar echo can be modeled as follows:

$$E(f,\varphi) = \iint \sigma(p_x, p_y) \exp\left\{ j\frac{4\pi f}{c}\left(p_x\cos\varphi + p_y\sin\varphi\right)\right\} dp_x dp_y \tag{4}$$

where $E(f,\varphi)$ is the radar echo, f is frequency of the electromagnetic wave, φ is the angle of radar observation, and c is the speed of light.

The basic principle of classical SAR imaging is employing Fast Fourier Transform (FFT) to reconstruct $\sigma(p_x, p_y)$. It is worth to mention that the actual radar echoes contain phase errors caused by non-ideal motion of target. These phase errors should be compensated before SAR imaging (Bao, 2006).

In this paper, we utilize CS for radar sampling and imaging. Firstly, under the hypothesis of "point scattering model" (Huang, 2006), the radar images are sparse in their original (space) domain, so we can take identity matrix as the sparse representation matrix.

Secondly, we will re-format the generation of radar echo in the framework of CS. Supposing the scattering coefficients $\sigma(p_x, p_y)$ can be represented by a $\sqrt{N}\times\sqrt{N}$ -dimensional matrix, and can be reshaped into a $N\times1$ -dimensional vector $\boldsymbol{\sigma}$. Matirx $\mathbf{F}$ stands for a $N \times N$ -dimensional Kronecker product matrix, which is produced by two identical $\sqrt{N}\times\sqrt{N}$ -dimensional Fourier transform matrices. Based on these, a discrete version of Eq. (4) can be described as

$$\mathbf{E} = \mathbf{F}\cdot\boldsymbol{\sigma}. \tag{5}$$

Where $\mathbf{E}$ is a matrix represents radar echoes. Furthermore, assuming an random index set $\boldsymbol{\Gamma} \subset \{1,\ldots,N\}$ obeys $|\mathbf{T}| = M$ □ N, then, a sub-matrix $\mathbf{F}_\Gamma$ can be formed by selecting M rows from $\mathbf{F}$ according to $\boldsymbol{\Gamma}$. So the compressive radar sampling process can be described as

$$\mathbf{E}_\Gamma = \mathbf{F}_\Gamma \cdot \boldsymbol{\sigma}. \tag{6}$$

Note that the dimension of $\mathbf{E}_\Gamma$ is far lower than $\mathbf{E}$.

Finally, the scattering coefficients vector $\boldsymbol{\sigma}$ is retrieved by sparse recovery as Eq. (3).

In our simulation, electromagnetic scattering computing software is employed to generate radar echoes of airplane A10. The overview of the 3D model of airplane A10 is shown in Fig. 1, and some radar parameters are list in Table 2.

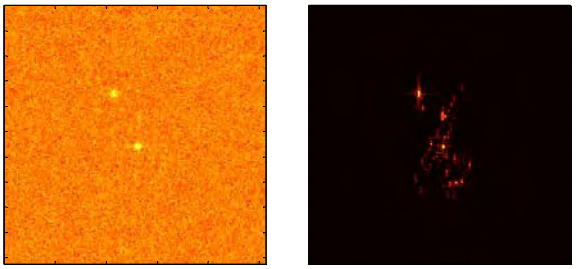
The sensing matrix $\mathbf{F}_\Gamma$ is constructed by randomly selecting 30% rows of the 65536×65536 dimensional Kronecker product matrix $\mathbf{F}$, and compressive measurement is then generated. The imaging results from conventional FFT-based and sparse recovery methods are listed in Fig. 3. We can see from this figure that there are only two intense scattering points in the first image and other scattering points are missed. The sparse recovered image (Fig. 3(b)) clearly describes the outline of airplane. This comparison can validate the feasibility of the new principle for radar sampling and imaging based on CS.

Besides, the resolution of SAR image also can be enhanced via sparse recovery. The resolution of conventional correlation-based imaging is limited by the ambiguity function of the transmitted signal; while the resolution of the sparse recovery is determined by the accuracy of optimization and the discretization of scattering coefficients $\boldsymbol{\sigma}$.

Figure 1. An overview of the 3D model of A10

Table 2. Radar observation parameters

Carrier frequency	9G Hz (X band)
Bandwidth	200M Hz
Azimuth angles	44°to 47°
Polar angles	60°
Amount of samples	256× 256



(a) (b)

Figure 3. Result comparing of the two methods based on 30% measured data. (a) conventional FFT-based method, (b) sparse recovery.

4. OPTICAL IMAGING BASED ON CS

The basic composition of the proposed random phase modulation and sparse sampling based imaging system is sketched in Fig. 4. The incident lights are firstly transformed to their frequency domain by a Fourier Transform (FT) lens. Then, the transformed lights will be leaded to a Spatial Light Modulator (SLM), and the SLM will add random phases to the lights which pass through its pixels. That is called random phase modulation in this paper. Subsequently, the modulated lights will be transformed back to space domain by an inverse Fourier Transform (IFT) lens. After that, the lights will be randomly and sparsely sampled by a two-dimensional detector array. A typical SLM consist of liquid crystal pixels, each independently addressed, acting as separate variable amplitude and phase modulator. In the proposed imaging system, the SLM is placed at the image plane of the FT lens.

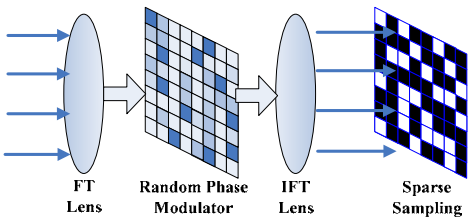


Figure 4. Schematics of the random phase modulation and sparse sampling system.

Suppose $\mathbf{x}_0$ is the reshaped $N\times 1$ -dimensional image (whose original dimension is $\sqrt{N}\times\sqrt{N}$), and its frequency spectrum is $\mathbf{F}\mathbf{x}_0$, where matrix is the same as defined in Eq. (5). The SLM can be mathematically described as a $N\times N$ diagonal matrix $\mathbf{\Sigma}$, whose non-zero entries are $\exp(-j\pi\cdot\varphi_n)$, $1\le n\le N$, where φ_n Uniform$([-1,1])$. The sparse sampling is realized by a multiplication of a $M\times N$ matrix $\mathbf{S}$, whose rows are randomly selected from a $N\times N$ diagonal matrix whose entries in diagonal line obey (0, 1) binominal distribution. Then, the whole process of Fig. 4 can be mathematically described as

$$\mathbf{y}=\mathbf{S}\mathbf{F}^{-1}\mathbf{\Sigma}\mathbf{F}\mathbf{x}_0 \qquad (7)$$

where $\mathbf{F}^{-1}$ stands for two-dimensional IFT matrix.

The function of the phase modulation (i.e. multiplication of matrix $\mathbf{\Sigma}$) is to translate phases of the spectrum to “noise like” modalities. So, when the modulated spectrum is transformed back to space domain, its energy will evenly spread out of the entire image plane. This means that each sample from the detector will carry some mixed information about the original image. If sufficient measurements (which is still much less than N) are obtained, we can reconstruct the image according to Eq. (3). In other words, the random phase modulation extended the space-bandwidth product of original signal. It, together with the sparse recovery, enables the system to acquire more detailed (high-resolution) images with fewer measurements.

The sparse representation in optical imaging is more complicated than radar imaging, where identity matrix is chosen. Borrowing the idea form JPEG and JPEG2000 stands, we employ Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT) to construct our sparse representation matrix.

The feasibility of the proposed random demodulation and spare sampling based optical imaging system is validated by a numerical experiment in this section. The original image in the simulation is show in Fig. 5(a). The measurements are generated according to Eq. (6), and $M/N\approx 0.5$. Fig. 5(b) shows an image from direct reconstruction, which is an inverse process of Eq. (6). The sparse recovery gives much better results. Two different images are shown in Fig. (c) and (d), which corresponding to DCT and DWT representation matrix respectively.

One of the superiorities of CS is reducing the measurement for image reconstruction. Fig. 6 shows the curve of RMSE of sparse recovery with DCT and DWT due to the variation of M/N .

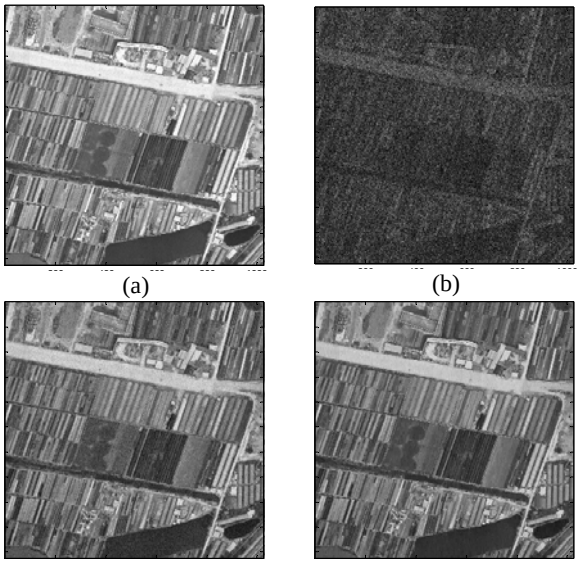


Figure 5. Result of the CS based optical imaging simulation. (a) original image, (b) direct reconstruction (RMSE: 0.30), (c) sparse recovery with DCT matrix (RMSE: 0.06), (d) sparse recovery with DWT matrix (RMSE: 0.05).

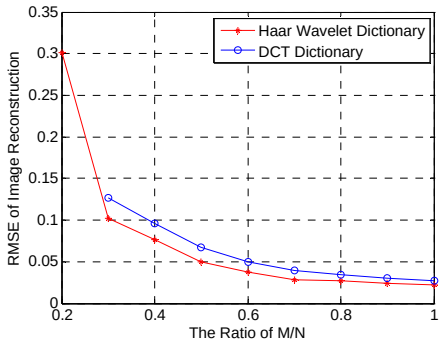


Figure 6. Curve of RMSE of sparse recovery with DCT and DWT due to the variation of M/N.

5. CONCLUSIONS

Benefit from its potential for alleviating the data sampling and storage system, CS theory has received more and more attentions. As an application in remote sensing area, the theoretical framework of SAR and optical imaging based on compressive sampling and sparse recovery is investigated in this paper. Numerical simulation validated the feasibility of the systems.

CS theory also can be employed in any signal acquisition system which can be re-formulated as an inversion of linear equations. However, the primary restriction is the computational efficiency in sparse recovery. Although we have emphasized this problem in our algorithm, the consumption of memory and time still substantially exceeds that of the conventional imaging method, particularly when the images are large.

Future research will focus on the further alleviation of the computational burdens. The other important point is the optimization of sensing matrix, which will permit exactly reconstruction with fewer measurements. The difficulty of physical realization will also be taken into account during the optimization.

REFERENCES

Baraniuk, R., and Steeghs, P., 2007. Compressive radar imaging, in Proc. *IEEE Radar Conference, Boston, MA*, pp. 128-133.

Bao, Z., Xing, M., and Wang, T., 2006. *Radar Imaging Technology*, Publishing house of electronics industry, Beijing. [in Chinese]

Cai, J., Osher, S., and Shen, Z., 2008. Linearized Bregman iterations for compressed sensing, *UCLA CAM Report* 08-06.

Candes, E., Romberg, J., and Tao, T., 2006. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, *IEEE Trans. Inform. Theory*, Vol. 52, pp. 489-509.

Candes, E., Romberg, J., and Tao, T., 2006b. Stable signal recovery from incomplete and inaccurate measurements, *Comm. on Pure and Applied Math.*, Vol. 59, pp. 1207-1223.

Candes, E., and Tao, T., 2006c. Near-optimal signal recovery from random projections and universal encoding strategies?, *IEEE Trans. Inform. Theory*, Vol. 52, pp. 5406-5245.

Chen, S., Donoho, D., and M. Saunders, 1999. Atomic decomposition by basis pursuit, *SIAM Journal on Scientific Computing*, Vol. 20, pp. 33-61.

Donoho, D., 2006. Compressed sensing, *IEEE Trans. Inform. Theory*, Vol. 52, pp. 1289~1306.

Duarte, M., Davenport, M., Takhar, D., Laska, J., Sun, T., Kelly, and K., Baraniuk, R., 2008 Single-pixel imaging via compressive sampling, *IEEE Signal Process. Mag.* Vol. 25, pp. 83-91.

Figueiredo, A., Nowak, R., and Wright, S., 2007. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems, *IEEE Journal of Selected Topics in Signal Processing*, Vol. 1, pp. 586-597.

Hale, E., Yin, W., and Zhang, Y., 2008. Fixed-point continuation for l1-minimization: Methodology and convergence, *SIAM J. on Optimization*, Vol. 19, pp. 1107-1130.

Huang, P., Yin, H., and Xu, Xi., 2006. *Radar Target's Characteristic*, Publishing house of electronics industry, Beijing. [in Chinese]

Osher, A., Mao, Y., Dong, B., and Yin, W., 2008. Fast Linearized Bregman iterations for compressed sensing and sparse denoising, *UCLA CAM Report* 08-37.

Romberg, J., 2008. Compressive Sensing By Random Convolution, http://dsp.rice.edu/files/cs/RandomConvolution-.pdf.

Stern, A., and Javidi, B., 2007. Random Projections Imaging With Extended Space-Bandwidth Product, *Journal of Display Technology*, Vol. 3, pp. 315~320.

Van Den Berg, E., and Friedlander, M., 2008. Probing the Pareto frontier for basis pursuit solutions, *SIAM Journal on Scientic Computing*, Vol. 31, pp. 890-912.

Wright, S., Nowak, R., and Figueiredo, M., 2009. Sparse reconstruction by separable approximation, *IEEE Transactions on Signal Processing* (to appear).

Wen, Z., Yin, W., Goldfarb, D., and Zhang, Y., 2008. A fast algorithm for sparse reconstruction based on shrinkage, subspace optimization and continuation, *Rice University CAAM Technical Report* TR09-01.

COMPARISON OF TERRESTRIAL LASER SCANNER AND SYNTHETIC APERTURE RADAR DATA IN THE STUDY OF FOREST DEFOLIATION

S. Kaasalainen [a, *], J. Hyyppä [a], M. Karjalainen [a], A. Krooks [a], P. Lyytikäinen-Saarenmaa [b], M. Holopainen [c], A. Jaakkola [a]

[a] Finnish Geodetic Institute, Department of remote sensing and photogrammetry, Finland – Sanna.Kaasalainen@fgi.fi
[b] University of Helsinki, Department of Forest Ecology, Finland
[c] University of Helsinki, Department of Forest Resource Management, Finland

KEY WORDS: TLS, LIDAR, SAR, Comparison, Forestry, Change Detection

ABSTRACT:

We present new results from a terrestrial laser scanner (TLS) based investigation on forest defoliation caused by the European pine sawfly (*Neodiprion sertifer*) in a Scots pine (*Pinus sylvestris*) dominated forest. The TLS results are compared with simultaneous ERS-2 Synthetic Aperture Radar (SAR) images in order to investigate the synchronous use of ALS and radar in forest change detection, and search for ground based validation methods for satellite SAR forest monitoring. The TLS and SAR based change detection is compared with visual estimation of the defoliation intensity as percentage of needle loss in the living crown. The agreement in results points out the potential for a combined method. The capability of TLS of deriving changes on the standing tree biomass and defoliation level was also confirmed by destructive, consecutive defoliation operations in laboratory. The biomass of the tree was measured simultaneously with the TLS measurements. The point cloud agreed with standing biomass with 95-99% coefficient of determination implying that terrestrial laser can be a powerful tool for biomass change reporting, and thus, usable for defoliation measurement.

1. INTRODUCTION

1.1 Remote Sensing of Forest Biomass

The international interest in biomass detection is strongly related to forest health and carbon cycle monitoring (Sexton et al., 2009). The need for improved tools for, e.g., carbon monitoring applications, and the shortage of data for accurate biosphere and climate models has been internationally recognized (e.g., Houghton et al., 2009). As the current knowledge on biomass, and particularly the changes related to it are almost entirely based on subjective ground measurements, remote sensing methods are called for. Airborne laser scanning (ALS) is a promising method for biomass detection because of its capability of direct measurement of vegetation structure and stand attributes (Næsset, 2002, Hyyppä and Inkinen, 1999). There are several recent activities and development of methods towards more accurate ALS-based biomass detection (Sohlberg et al., 2006; Zhao et al., 2009; Hawbaker et al., 2009).

1.2 Terrestrial Laser Scanning in Forest Remote Sensing

The environmental applications of terrestrial laser scanner (TLS) are constantly increasing. TLS has been found to be an effective and low-cost monitoring method, and the information on TLS performance and range data accuracy is constantly increasing (see Kaasalainen et al., 2009 for more references). The number of TLS applications is increasing in forest management and agriculture, e.g., in measuring the 3D-structure of trees and vegetation canopies (canopy gap fractions), tree volumes and leaf-area. TLS has been used for modelling of individual trees and canopies in (e.g., Pfeifer et al., 2004; Pfeifer and Winterhalder, 2004; Gorte and Pfeifer, 2004; Hosoi and Omasa, 2006; Fleck et al., 2007; Danson et al., 2007; Xu et al., 2007; Chasmer et al., 2006), and for automatic forest parameter estimation (e.g., Bienert et al 2006a, 2006b, 2007; Aschoff and Spiecker, 2004; Király and Brolly, 2007). The use of TLS has recently been extended into mobile methods, namely vehicle-based laser scanning, because of their efficiency in producing large amounts of high-resolution data and strong future potential for different applications (Jaakkola et al., 2008).

1.3 SAR-based Forest Remote Sensing

A large number of SAR-based methods have been created for detection of forest parameters, such as canopy height (Garestier et al., 2008; Sexton et al., 2009). E.g., airborne polarimetric SAR has been found a promising method for estimating the forest structure and tree height (Garestier et al., 2009). In comparison with radar, LiDAR, and field measurements, Sexton et al. (2009) found that the LiDAR measurement produced the best accuracy in pine forests, while the interferometric (SAR) showed potential for LiDAR based calibration. In their investigations of radar-LiDAR synergy, Nelson et al. (2007) emphasized the efficiency of LiDAR in producing accurate and precise biomass estimation and improving the accuracy of radar data. While laser scanning produces higher resolution and accuracy, radar methods are less weather-dependent and capable of producing data from large areas with high temporal resolution (Holopainen et al., 2009b). Holopainen et al. (2009a) compared E-SAR, Landsat Extended Thematic Mapper (ETM) and aerial photographs in estimation of plot-level forest variables and reported relative root-mean-squared-errors (RMSEs) for E-SAR of 45%, 29%, 28% and 38% for volume (m3/ha), mean diameter (Dg; cm) , mean height (Hg; m) and basal area (BA; m2/ha), respectively. In combining E-SAR with aerial photographs, the relative RMSEs for the same variables were 38%, 26%, 23% and 33%.

This article presents first results from a TLS-based investigation on forest defoliation caused by the European pine sawfly, specialist feeder consuming mature needles of pine trees early in

* Corresponding author.

the season (Viitasaari and Varama, 1987). The TLS results are compared with simultaneous ERS-2 SAR images in order to investigate the synchronous use of LiDAR and radar in forest change detection, and search for validation methods for SAR. SAR data are available from large areas and better temporal resolution than laser data, but because of their higher accuracy, laser-based methods would provide an efficient validation for SAR change detection and forest parameter estimation. This requires a systematic comparison of simultaneous data from both sensors.

2. METHODS

2.1 The Study Site and the Pine Sawfly Hazard

This study was a part of ongoing monitoring campaign for forest defoliation caused by pine sawflies (Diprionidae). The traditional monitoring methods have been based on field sampling (e.g., manual collection of different life stages) and subjective visual observation of tree condition. The test site was located in Outokumpu, eastern Finland, in a Scots pine dominated forest. Reference data were collected from 20 field plots (consisting of 526 trees in total) in June 6-9, 2009. The visual assessment of defoliation was carried out simultaneously with tree-wise measurements in the field plots, and an additional visual assessment was done after defoliation by sawfly larvae (July 26-28, 2009) (see Lyytikäinen-Saarenmaa et al., 2006; De Somviele et al. 2007 for more details).

2.2 Terrestrial Laser Scanning

The terrestrial laser scanner used in this study was Leica HDS6000, a 685nm phase-based continuous wave laser scanner with a 360°×310° field-of-view. The distance measurement accuracy is 4-5mm and the angular resolution is selectable from full 0.009° down to 0.288°. The circular beam diameter at the exit and the beam divergence are 3 mm and 0.22 mrad, respectively. The scanner uses a silicon Avalanche Photo Diode (APD) as a photo detector.

Two sets of TLS measurements were carried out in the study site in June 25 and July 26-27 during the active period of the pine sawfly hazard. The first measurement was made at the early phases of defoliation and the second one after defoliation period. The scanning was performed from the centre of each field plot with similar scanning parameters and resolution at both dates. Directly visible trees (i.e., those not obscured by other trees) were extracted from the resulting laser point clouds, and the change in the number of laser returns from each tree was compared to the visually estimated defoliation intensity. The defoliation intensity of trees at each field plot was visually assessed and expressed as a relative percentage of needle loss compared with a reference, imaginary tree with full, healthy foliage. The difference from the reference was expressed in incremental defoliation classes of 10%. For full details of this standardized method, see Eichhorn (1998).

The capability of TLS of deriving changes on the standing tree biomass and defoliation level was verified by destructive, consecutive defoliation operations by two different laboratory case studies. In the first case study, one Scots pine tree was defoliated (by picking the needles) in the laboratory in 7 steps. The biomass of the tree (i.e., the mass of the picked needles, and in the end also the stem and branches) was measured simultaneously with the TLS (the scanner located beside the trees), and defoliation was also estimated visually. Three simple tree-wise parameters were derived to represent the tree quality and defoliation: number of echoes coming from the tree, ratio of tree hits (i.e., number of echoes coming from the tree divided by the total number of echoes), and the number of echoes from the ground. These parameters were compared with the tree biomass (in grams) measured with a 2-g accuracy for each step.

In the second laboratory experiment, five pine (Scots Pine) and spruce (Norway Spruce, *Picea abies*) trees were measured with TLS from above and below. Here too the biomass of the tree was measured simultaneously with 2-g accuracy. The biomass was measured in five steps for each tree resulting in 50 samples of biomass/defoliation.

2.3 SAR Image Analysis

The Outokumpu research site was also investigated from the ERS-2 SAR images. ERS-2 is an Earth observation satellite of the European Space Agency launched in 1995. One of its main instrument is SAR, which uses C band microwave radiation (wavelength of about 5.6 cm). The spatial resolution of SAR in the image mode (PRI) is about 25 metres in range and 21 metres in azimuth direction. Some of the main functions of the satellite have damaged preventing for example the interferometric processing in some cases, but the radiometric stability of the PRI images is still expected to be fine (Meadows et al., 2008). The reason for using ERS-2 images was their availability in the ESA image archive.

Satellite	Pass	Track	Frame	Date
ERS-2	Descending	93	2336	11 Aug 2008
ERS-2	Descending	93	2336	15 Sep 2008
ERS-2	Descending	93	2336	20 Oct 2008
ERS-2	Descending	93	2336	18 May 2009
ERS-2	Descending	93	2336	22 Jun 2009
ERS-2	Descending	93	2336	27 Jul 2009
ERS-2	Descending	93	2336	31 Aug 2009

Table 1. List of used ERS-2 SAR images.

Altogether 7 images were acquired through the ESA Category-1 project (Table 1). Topography is known to have a strong effect on the observed backscattering. Therefore, we decided to use images with the same imaging geometry (same satellite pass direction and track). However, the weather conditions during the acquisition, as well as the soil surface and vegetation moisture values of the test plot, are not known. Therefore, it can be expected that there is variation in the backscattering values of the test plots between the images. All images were ordered as detected products (PRI images), so only the amplitude information of backscattering signal is used in the studies (non-interferometric data).

First, all input images were co-registered with each other with sub-pixel accuracy. The image of 11 August 2008 was used as a master image. Then, the stack of images was georeferenced into the Finnish map coordinate system (ETRS-TM35FIN) using Ground control points and Digital elevation model. According to the image coordinate residuals the accuracies of 0.8 pixels and 0.4 pixels were achieved in the range and azimuth coordinate directions, respectively. A false colour fusion of all input images is presented in Figure 2, where blue areas correspond to the water bodies. The city of Outokumpu is located in the bottom right corner of the Figure 2.

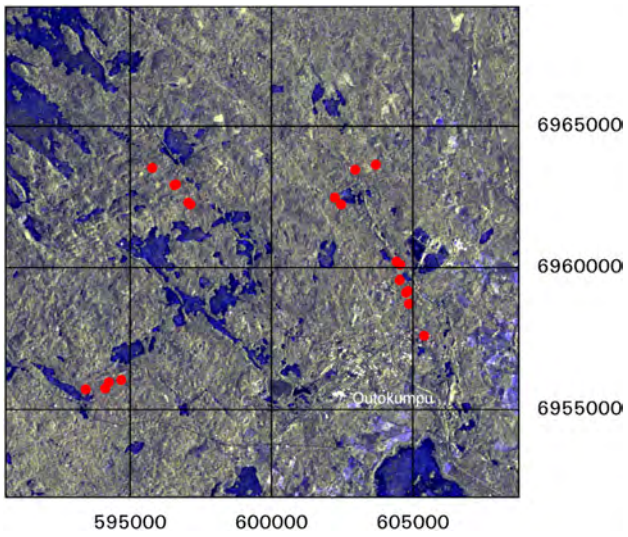


Figure 2. Fusion of all input SAR-2 images. Red dots present the locations of the test plots (Red+Green: Mean amplitude of all images, Blue: Standard deviation of amplitude in input images). Map coordinate system: UTM35N/WGS84. Original data © 2008-2009, European Space Agency.

3. RESULTS

3.1 TLS Laboratory Measurements of Biomass Change and Defoliation

In the first laboratory measurement with the used three predictors, the Pearson correlation coefficients derived were 0.996, 0.977 and 0.929, respectively, which implies that the change in the number of points reflected from the tree represent accurately the change in the standing biomass of the tree. The correlation with the estimated visual defoliation was much smaller. Thus, the changes of the pine biomass could be accurately determined by the change of the laser points in multi-temporal TLS surveys. In this test the reduction of needles from the tree was linear to the reduction of the hits from the tree. In practice, we expect that the process could be non-linear or it could depend on tree species and density of the trees.

In the second laboratory experiment consisting of 10 trees (five pines and five spruces), we used the relative number of hits coming from each tree as a predictor (normalized into 1 when no defoliation has occurred) as learnt from the first experiment. The TLS position (two scanners, one above and the other from below) did not affect the results neither did the tree species. The coefficient of determination with a linear regression model was 0.95 for total biomass and 0.97 for needle and branch biomass (Figures 3 and 4), which implies that the relative number of points reflected from the tree represent accurately the relative standing biomass of the tree. There was some non-linearity in the response, especially with lower relative number of hits, and with a non-linear model the results could be improved.

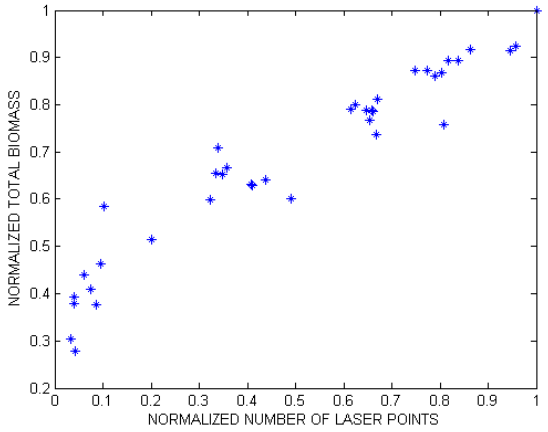


Figure 3. Laboratory experiment: normalized laser point number versus normalized total biomass of the trees. R^2=0.95.

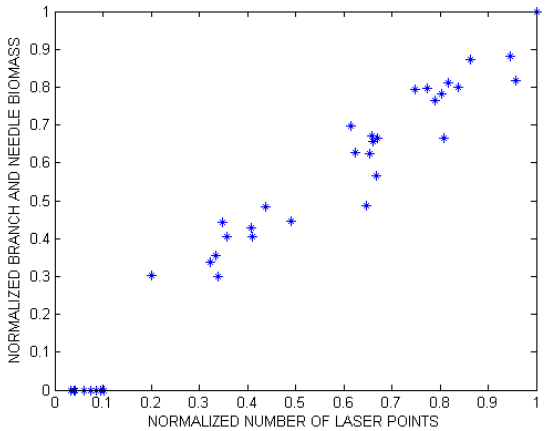


Figure 4. Normalized laser point number versus normalized branch and needle biomass of the trees. R^2=0.97.

3.2 TLS Field Experiment

The change in laser returns from the tree canopy (in percentages) between the measurements in June 25 and July 26-27 is presented in Figure 5 and compared with visual observation. Clear trends of defoliation are visible in both visual and TLS-based analysis of the tree canopies. There are several factors that affect the accuracy, e.g.,

- Deviation in change detection data from TLS point clouds (e.g. mixed pixel effect caused by the phase-based scanning technique)
- The first visual estimation was carried out in June 6-9, i.e. two weeks before the first TLS measurements. During the time difference, the youngest larval instars were already consuming needles.
- The visual estimation of defoliation is subjective (compared to, e.g., laboratory measurements, for which the accuracy is better, see Sect. 3.1)
- The ground-based TLS measurement is made in the upward direction, preventing some parts of the canopy from being measured. Measurement facing downwards (e.g., ALS or unmanned aerial vehicle (UAV) based TLS) would improve the accuracy.

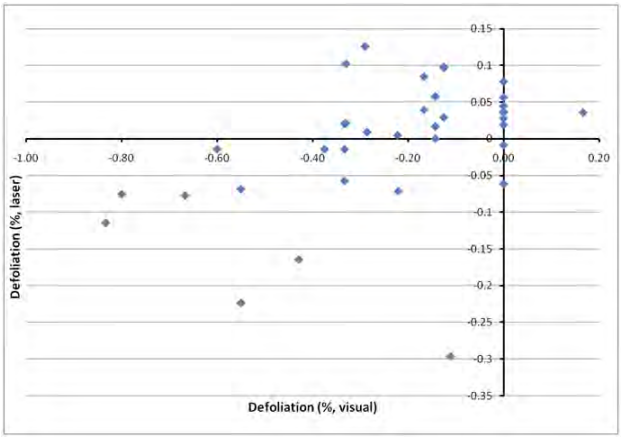


Figure 5. Visual and TLS-based measurement of forest defoliation (in percentages).

3.3 SAR

The change in the ERS-2 backscattering values (amplitude values) of the test plots in comparison with defoliation intensity in late July 2009 is presented in Figure 6. The backscattering values were averaged from the SAR pixel values using a circle with the radius of 50 metres. The test plots were located in the centre of the circles. It should be noted that the spatial and temporal pattern of defoliation, i.e., the focal point and the damage intensity change each year, so the defoliation is not uniform. Therefore, averaging strongly generalizes the results concerning the defoliation. However, averaging was the only way to reduce SAR speckle, and therefore this result is preliminary.

According to the results of 19 test plots, a slight change in the averaged SAR backscatter occurred for the plots with visually observed defoliation. Little or no change in backscatter was observed for those plots where (visual) signs of defoliation were not present. Nevertheless, more test plots for ground reference and SAR data are needed to find out the role of possible other factors to these changes.

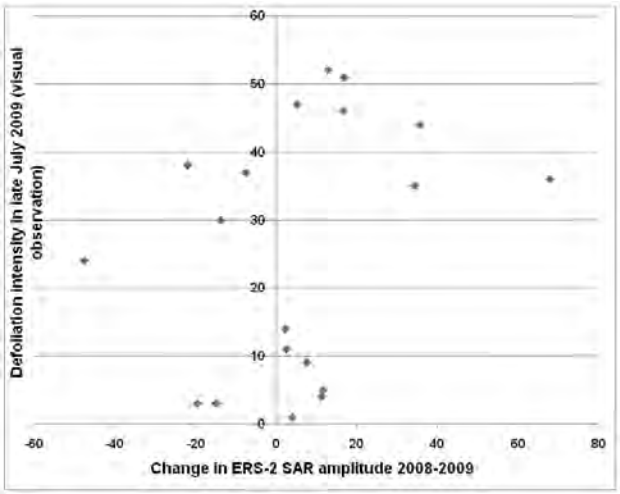


Figure 6. Scatterplot of defoliation intensity (in percentages) in 2009 and change in ERS-2 backscattering (2008-2009).

Because of the qualitative nature of these first results, comparison between TLS and SAR is difficult, but the similarity in results suggests that combined use of these two methods would provide the best time resolution in long-term monitoring applications. In their earlier study, Holopainen et al (2010) compared the accuracy of ALS, multi-temporal high-resolution non-interferometric TerraSAR-X (TSX) radar data and combined feature set derived from these data in the estimation of forest variables (mean volume, basal area, mean height, mean diameter and tree species-specific mean volumes) at plot level. The combined feature set marginally outperformed the ALS-based feature set. Features from TSX alone performed poorly, but brought some extra information into the combined set. However, due to favourable temporal resolution, they concluded that satellite-borne radar imaging is a promising data source for updating large-area forest inventories based on ALS.

CONCLUSIONS

The defoliation trend was visible in the TLS laboratory and field experiments, and SAR data for those plots where defoliation was observed visually. Based on these results and the earlier similar findings (Holopainen et al., 2010), there is great potential in synchronized use of laser scanner and SAR in change detection and forest parameter estimation. This requires careful planning and timing of all experiments for simultaneous data from both sensors and ground reference.

Future tests will show whether the effect of the trunk is important. The results partly explain why ALS has been very effective in earlier studies for stem volume estimation. ALS does not measure only height information, but also direct metrics for biomass, or at least for biomass change. Future tests are needed to verify this, but our results integrated with earlier findings propose that laser scanner measures directly tree height, crown area and biomass with a relatively high accuracy.

The laboratory results point out the better accuracy of TLS measurements, but SAR data are available from large areas and better temporal resolution than laser data. Therefore a combined approach would often produce a larger coverage of observations both spatially and temporally. Also, a shorter wavelength in SAR (X-band, 3.1 cm, as in TerraSAR-X or Cosmo-SkyMed) might be better than the C-band (5.6 cm) used in this experiment.

REFERENCES

Aschoff, T., and Spiecker, H., 2004. Algorithms for the automatic detection of trees in laser scanner data. In: *IAPRS working group VIII/2 "Laser-Scanners for Forest and Landscape Assessment"*, University of Freiburg, pp. 66-70.

Bienert, A., Scheller, S., Kesane, E., Mullooly, G. and Mohan, F., 2006a. Application of terrestrial laser scanners for the determination of forest inventory parameters. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Dresden, Germany, Vol. 36, Part 5.

Bienert, A., Maas, H., and Scheller, S., 2006b. Analysis of the information content of terrestrial laserscanner point clouds for the automatic determination of forest inventory parameters. In: *ISPRS WG VIII/11 & EARSeL joint Conference '3D Remote Sensing in Forestry'*, Vienna, 14-15 February.

Bienert, A., Scheller, S., Keane, E., Mohan, F. and Nugent, C., 2007. Tree detection and diameter estimation by analysis of

forest terrestrial laserscanner point clouds. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Espoo, Finland, Vol. 36, Part 3, pp.50-55.

Chasmer, L., Hopkinson, C. and Treitz, P., 2006. Investigating laser pulse penetration through a conifer canopy by integrating airborne and terrestrial lidar. *Canadian Journal of Remote Sensing*, 32(2), pp. 116-125.

Danson, F.M., Hetherington, D., Morsdorf, F., Koetz, B., and Allgöwer, B. Forest canopy gap fraction from terrestrial laser scanning. *IEEE Trans. Geosci. Remote Sensing*, 4(1), pp. 157-160.

Eichhorn, J., 1998. *Manual on Methods and Criteria for Harmonized Sampling, Assessment, Monitoring and Analysis of the Effects of Air Pollution on Forests. Part II. Visual Assessment of Condition and Submanual on Visual Assessment of Crown Condition on Intensive Monitoring Plots.* United Nations Economic Commission for Europe Convention on Long-Range Transboundary Air Pollution. Germany.

Fleck, S., Obertreiber, N., Schmidt, I., Brauns, M., Jungkunst, H.F., and Leuschner, C., 2007. Terrestrial lidar measurements for analysing canopy structure in an old-growth forest. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Espoo, Finland, Vol. 36 Part 3/W52, pp. 125-129.

Garestier, F. Dubois-Fernandez, P.C., and Papathanassiou, K.P., 2008. Pine forest height inversion using single-pass X-band PolInSAR data. *IEEE Trans. Geosci. Remote Sensing*, 46(1), pp. 59-68.

Garestier, F., Dubois-Fernandez, P.C., Guyon, D., and Le Toan, T., 2009. Forest Biophysical Parameter Estimation Using L- and P-Band Polarimetric SAR Data. *IEEE Trans. Geosci. Remote Sensing*, 47(10), pp. 3379-3388.

Gorte, B., and Pfeifer N., 2004, Structuring laser-scanned trees using 3D mathematical morphology. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Istanbul, Turkey, Vol. 35, Part B5, pp. 929-933.

Hawbaker, T.J., Keuler, N. S., Lesak, A.A., Gobakken, T., Contrucci, K., and Radeloff V.C., 2009. Improved estimates of forest vegetation structure and biomass with a LiDAR-optimized sampling design. *J. Geophys. Res.*, 114, G00E04.

Holopainen, M., Tuominen, S., Karjalainen, M., Hyyppä, J., Vastaranta, M. and Hyyppä, H. 2009a. The accuracy of high-resolution radar images in the estimation of plot-level forest variables. In: Sester, M., Bernard, L., and Paelke, V.: *Advances in GIScience. Lecture notes in geoinformation and cartography.* Springer, pp. 67-82.

Holopainen, M., Haapanen, R., Karjalainen, M., Vastaranta, M., Hyyppä, J., Yu, X., Tuominen, S., and Hyyppä, H., 2009b. Combination of low-pulse ALS data and TerraSAR-X radar images in the estimation of plot-level forest variables. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Paris, France, Vol. 38, Part 3/W8, pp. 135-140.

Holopainen, M., Haapanen, R., Karjalainen, M., Vastaranta, M., Hyyppä, J., Yu, X., Tuominen, S., and Hyyppä, H. 2010. Comparing accuracy of airborne laser scanning and TerrSAR-X radar images in the estimation of plot-level forest variables. *Remote Sensing*, 2(2), pp. 432-445.

Hosoi, F, and Omasa, K., 2006. Voxel-based 3-D modeling of individual trees for estimating leaf area density using high-resolution portable scanning lidar. *IEEE Trans. Geosci. Remote Sensing*, 44(12), pp. 3610-3618.

Houghton, R.A., Hall, F., and Goetz, S.J., 2009. Importance of biomass in the global carbon cycle. *J. Geophys. Res.*, 114, G00E03.

Hyyppä, J., and Inkinen, M., 1999. Detecting and estimating attributes for single trees using laser scanner. *The Photogrammetric Journal of Finland*, 16, pp. 27-42.

Jaakkola, A., Hyyppä, J., Hyyppä, H., Kukko, A., 2008. Retrieval Algorithms for Road Surface Modelling Using Laser-Based Mobile Mapping. *Sensors*, 8(9), pp. 5238-5249.

Kaasalainen, S., Krooks, A., Kukko, A., and Kaartinen, H., 2009. Radiometric Calibration of Terrestrial Laser Scanners with External Reference Targets. *Remote Sensing*, 1(3), pp. 144-158.

Király, G., and Brolly, G., 2007. Tree height estimation methods for terrestrial laser scanning in a forest reserve. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Espoo, Finland, Vol. 36, Part 3, pp. 211-215.

Lyytikäinen-Saarenmaa, P., Varama, M., Anderbrant, O., Kukkola, M., Kokkonen, A.-M., Henderström, E., and Högberg, H.-E., 2006. Monitoring the European pine sawfly in maturing Scots pine stands with pheromone traps. *Agricultural and Forest Entomology*, 8, pp. 7-15.

Meadows, P., Miranda, N., Pilgrim, A., Tranfaglia, M., 2008. The Performance of the ERS-2 SAR on becoming a Teenager, In: *Proceedings of the CEOS SAR Cal/Val Workshop*, Oberpfaffenhofen, Germany, 27-28 November 2008.

Nelson, R.F., Hyde, P., Johnson, P., Emessiene, B., Imhoff, M.L., Campbell, R., and Edwards, W., 2007. Investigating RaDAR–LiDAR synergy in a North Carolina pine forest. . *Remote Sens. Environ.*, 110(1), pp. 98-108.

Naesset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sens. Environ.*, 80, pp. 88-99.

Pfeifer, N., Gorte, B., and Winterhalder D., 2004. Automatic reconstruction of single trees from terrestrial laser scanner data. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Istanbul, Turkey, Vol. 35, Part B5, pp. 114-119.

Pfeifer, N. and Winterhalder D., 2004. Modelling of Tree Cross Sections from Terrestrial Laser-Scanning Data with Free-Form Curves. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol 36, Part 8/W2, pp. 76-81.

Sexton, J.O., Bax, T., Siqueira, P., Swenson, J. J., and Hensley, S., 2009. A comparison of lidar, radar, and field measurements of canopy height in pine and hardwood forests of southeastern North America. *Forest Ecology and Management* 257(3), pp. 1136-1147.

Sohlberg, S., Næsset, E., Hanssen, K.H., and Christiansen, E., 2006. Mapping defoliation during a severe insect attack on Scots pine using airborne laser scanning. *Remote Sens. Environ.*, 102, pp. 364-376.

De Somviele, B., Lyytikäinen-Saarenmaa, P. and Niemelä, P., 2007. Stand edge effects on distribution and condition of Diprionid sawflies. *Agricultural and Forest Entomology*, 9, pp. 17-30.

Viitasaari, M., and Varama, M., 1987. Sawflies 4. Conifer sawflies (Diprionidae). *University of Helsinki. Department of Agricultural and Forest Zoology. Reports*, 10, pp. 1-79. (In Finnish with an English summary)

Xu, H., Gosset, N., and Chen, B., 2007. Knowledge and heuristic-based modeling of laser scanned trees. *ACM Transactions on Graphics*, 26(4), Article 19.

Zhao, K., Popescu, S., and Nelson, R., 2009. Lidar remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers. *Remote Sens. Environ.*, 113(1), pp. 182-196.

ACKNOWLEDGEMENTS

The SAR data was related to the ESA project DUE Innovators II – Insect Combat (ESRIN/AO/1-5781/08/I-IC) and the authors want to thank Prof. Barbara Koch at Steinbeis Transferzentrum FELIS. This study was financially supported by the Academy of Finland (projects "New techniques in active remote sensing: hyperspectral laser in environmental change detection" and “Forest Tomography”). The study was also supported by the Academy of Finland project “Improving Forest Supply Chain by Means of Advanced Laser Measurements (L-Impact)”. The authors would like to thank Mikko Vastaranta at the University of Helsinki for participation in the experiments.

TRUE ORTHOPHOTO CREATION THROUGH FUSION OF LIDAR DERIVED DIGITAL SURFACE MODEL AND AERIAL PHOTOS

A. Kato [a, *], L. M. Moskal[b], P. Schiess[b], D. Calhoun[c], M. E. Swanson[d]

[a] Graduate School of Horticulture, Chiba University, 648 Matsudo Matsudo-shi Chiba 2718510 Japan
akiran@faculty.chiba-u.jp
[b] Precision Forestry Cooperative, School of Forestry, College of Environment, University of Washington, Box 352100 Seattle, WA 98195-2100 USA
(lmmoskal, schiess)@u.washington.edu
[c] Laboratoire d'Etudes des Transferts et de Mecanique des Fluides, Commissariat l'Energie Atomique F-91191 Gif-sur-Yvette Cedex, France
donna.calhoun@cea.fr
[d] Department of Natural Resource Sciences, Washington State University 177 Johnson Hall Pullman, WA 99164-6410 USA
markswanson@wsu.edu

KEY WORDS: Digital, LIDAR, Fusion, Orthorectification, Orthoimage

ABSTRACT:

Data fusion between aerial photos and LiDAR provides better estimates in forestry and ecological applications, because LiDAR mainly provides the structural information of objects and aerial photo can add spectral information to them. Without the data fusion, an accurate identification of tree crown information from two dimensional data is difficult due to shaded and shadow pixels cast on the image and image distortion. The aerial photogrammetric techniques cannot reconstruct the objects accurately in three dimensional spaces if they are not clearly visible on the photos. The conventional orthophotos, therefore, still have image distortion due to an inappropriate Digital Surface Model (DSM). LiDAR provides a more suitable surface of tree crown structure in three-dimensional spaces. This LiDAR-derived DSM could be used in conjunction with conventional photogrammetric techniques to rectify aerial photos and produce true orthophotos for each image. The existence of different perspective points from the use of multiple images results in different illumination and shadows cast on the DSM from the angle between the sun and the camera. In previous studies, a Z-buffer algorithm was applied for the occlusion detection and compensation. However, the technique was computationally intensive. In this study, the camera view and sun-oriented hillshade were generated using the LiDAR-derived DSM. The hillshade surfaces distinguished between the exposed and the occluded side of the DSM during the composition process of respective true orthophotos. This technique constituted a simpler approach and is applicable to data fusion between LiDAR and multispectral imagery to make an orthographically rectified image.

1. INTRODUCTION

1.1 Data Fusion

Data fusion data derived from different remote sensing sensors has been used for numerous applications. The improvement of detecting objects was demonstrated by data fusion among several different sensors to derive better results than that derived by each sensor solely. There are limits in two dimensional image analyses for forestry application. Those issues and the advantage of data fusion are discussed in the following subsection.

1.2 The limitation of two-dimensional image analysis

The limits of two dimensional image data have been addressed with several crown delineation techniques such as image processing algorithm (Brandtberg and Walter, 1998; Erikson, 2003), local maximum filter with semivariogram (Wulder et al., 2000), and valley following algorithms (Gougeon, 1995; Leckie et al, 2003, 2005). Brandtberg and Walter (1998) processed a near-infrared image to delineate tree crowns. They detected crown edges on the image for crown segmentation. The accuracy was 70% compared with manual delineation. Such delineation only worked for the visible portion of tree crowns because of the light illumination. Natural forests canopies are comprised of tree crowns with various sizes and shapes and complex vertical tree composition. This presents a challenge for crown delineation based on two-dimensional data. Leckie et al. (2005) found several other important issues in two-dimensional image analysis: different sun angles among multi-year data, different sensor view angles, and similar spectral signatures of trees within the old growth stand. The various shade and shadowing effects caused different delineation results. The sensor view angle, therefore, should be considered for further improvement of crown delineation technique.

1.3 Advantage of data fusion

For the conventional photogrammetric technique, creating automated Digital Surface Model (DSM) over a dense forest is prone to error because of the difficulty of matching pixels in the tree canopy between two stereo photos taken at different viewing angles. Light Detection and Ranging (LiDAR) is, therefore, a good alternative tool to create better DSMs to rectify the aerial images. The LiDAR derived Canopy Height Model (CHM) was used for the valley following algorithm to compare with the results from digital aerial photos (Leckie et al., 2003). They found that digital aerial photos were better for the delineation of tree crowns; while the LiDAR derived CHM was

* Corresponding author

better for open canopy areas. The LiDAR derived CHM was better suited in clearly distinct area among neighbouring pixels between trees and shrub (or ground). LiDAR was well suited to measure tree height and large tree crown delineation (Leckie et al., 2003). For the data fusion between aerial photos and LiDAR, a simply approach can be taken to form a direct link between LiDAR and orthophoto plainmatric coordinate. The orthophoto still has, however, distorted objects because of the insufficient quality of DTMs or DSMs that are generally used to rectify the images. Therefore, the color values assigned on an orthophoto does not generally match with the LiDAR coordinates. The orthophoto creation should be improved by better data fusion approaches.

As the better applications of data fusion between the LiDAR derived high resolution data and aerial photos, St-Onge et al (2004) used LiDAR derived Digital Terrain Model (DTM) to measure tree height using stereo pair of aerial photos. Their approach can be useful for the change detection of tree height over time when historical photos are applied for comparison. Rönnholm et al. (2004) used the backprojection approach with exterior orientation parameters derived by aerial triangulation process from stereo pair of aerial photos. They draped LiDAR points on terrestrial digital images. Their approach explicitly showed which part of tree crown the LiDAR points were reflected from and was used for calibration in LiDAR acquisition accuracy.

1.4 True orthophoto generation

A 'true' orthophoto has been created for detecting buildings in the urban area (Amhar et al., 1998, Habib et al., 2007, Rau et al., 2002, Schickler and Thorpe, 1998, Zhou et al., 2005) and forested area (Sheng et al., 2003). Ambar et al. (1998) defined "true orthophoto" as an orthophoto that is rectified orthographically. The true orthophoto, therefore, does not have any image distortion of objects on the final image. Ideally it rectifies the structure of trees in an upright position on the image. Ambar et al. used Digital Building Models (DBMs) to find the visible side of building from the photo projection center using Z-buffer algorithm. The Z-buffer algorithm is also called hidden surface removal in computer graphics (Angel, 2003) and determines the visibility of the objects on the photos. In transformation of ground coordinate to image space coordinate, the corresponding location between them is a many-to-one relationship (Sheng et al., 2003). The Z-buffer algorithm uses the distance from the photo projection center and determines the closest object to the projection center and occludes all others. In this way, the ground coordinates of the visible sides of objects are only selected and matched with the corresponding pixel on the photo. Furthermore, it was reported that the false occlusion was found using the direct method of Z-buffer algorithm, if the sampling distance on the image was less than the DSM resolution (Amhar et al., 1998, Habib et al., 2007, Rau et al., 2002). Habib et al. (2007) showed that false occlusion caused the black grid lines in the ground coordinate as an example when the sampling distance did not match between the image space and the ground coordinate. Habib er al. (2007) also used radial and spiral sweep methods to resolve the issue of the false occlusion without DBMs in urban area. The occluded area of a master image was filled with images from a slave image (Rau et al., 2002). In other words, orthophotos created from different camera view angles were used to compensate the occluded area to produce a gapless composite image (Sheng et al,, 2003). When the finer resolution true orthophotos were merged into a larger one, occluded areas were filled in by the color values derived from orthophotos using an angle-based photo composition scheme (Sheng et al., 2003), seam line adjustment (Schickler and Thorpe, 1998), and histogram matching (Rau et al., 2002, Zhou et al., 2005). These methods were complicated, because the color values from different perspective photos were variable and not well calibrated. There was a photographic "light fall-off effect" which creates the darker colors at the edge (Sheng et al., 2003). A relatively simpler approach using the hillshade function is, therefore, introduced in a large area for occlusion detection and compensation in this study.

1.5 Hillshade method for occlusion detection and compensation

Hillshade surface utilized by ArcGIS (ESRI Inc.) has been used for surface temperature change analysis (Hais and Kučera, 2008), making moisture index in landscape (Iverson et al., 1997), the visual investigation of landslide area (Van Den Eeckhaut et al., 2005, 2007), identifying tectonic phase (Pellegrini et al., 2003), and urban growth modeling (Mahiny and Gholamalifard, 2007). The hillshade function has been mainly used for landscape analysis and has not been used for true orthophoto generation. For forestry application, Kane et al. (2008) used the hillshade function for self-shadowing effect to characterize the stand condition. They cast the shadow artificially on the forested landscape using LiDAR derived CHM.

The advantage of this function is to distinguish which side of a surface is visible to the light source. The light source can be the sun or camera location. The parallel rays from the sun were used for hillshade when the shadow is cast onto the surface, because the distance of the sun is virtually infinite. But the camera location mounted on airborne vehicle is in lower altitude and this function was, therefore, customized to have radial arrays from camera location to cast the shadow on the surface (a central projection).

2. OBJECTIVES

In order to compose a LiDAR rectified image that is displacement free, this paper used photogrammetric relationships to assign color values from original perspective image directly to the LiDAR derived DSM (the backprojection process was used). The assigned color was mapped onto its LiDAR plainmatric coordinates in two dimensions to make a true orthophoto from each perspective image. At the final stage of the process, the light reflectance between camera view and sun angle was considered using hillshade function. The shadowing effect of hillshade surface, therefore, was used for occlusion detection and compensation during the composition process. The main discussion points of this paper are:

1) accuracy of aerial triangulation* and backprojection process.
2) occlusion detection and compensation using sun and camera view hillshade.

* aerial triangulation in this study is achieved by using the Leica Photogrammetric Suite (LPS, Leica Geosystems, Inc.).

In this paper, a LiDAR derived DSM is used to rectify each original perspective photo to make a true orthophoto. To make a gapless composite image, the occluded areas are detected using values of hillshade function.

3. DATA

The research area is located in the Washington State Park Arboretum at the south end of the University of Washington and east of downtown Seattle, WA. The total area is 0.93 km^2. The terrain of this study area is moderate in slope, and the site is in a relatively urban setting. Arboretum is in the city so that building and other distinctive objects are readily available for Ground Control Points (GCPs) over the site for the aerial triangulation process for the data fusion technique.

Aerial photos show an electromagnetic reflectance of solar radiation. The reflectance depends on the angle between sun and the projection center of each photo relative to the objects on and above the ground. These reflectance values are not calibrated like multispectral images. Each aerial photo collects one central projected perspective view. The camera settings used in this study are shown in Table 1. Three consecutive aerial photos (named Image1, Image2, and Image3 from the South) along the flight line are used with 60% overlap (endlap) between photos from south to north of the Arboretum. Note that the capture dates of the photos and LiDAR were different. The year was, however, the same. The solar horizontal and vertical angle are calculated by their timestamp using a solar position calculator provided by NOAA (National Oceanic and Atmospheric Administration) Surface Radiation Research Branch.

Airborne LiDAR data was acquired in the same coverage area of aerial photos at the same year. LiDAR sensor setting is shown in Table 2. Aerial photo was taken during leaf-on season and LiDAR was acquired during leaf-off season. Stands of this study area are heterogeneous, multi-aged mixtures of coniferous and deciduous tree species. The seasonal error can by caused over the deciduous tree area.

Camera of aerial photos	
Date of acquisition	July 26th, 2005
Camera type	Jena LMK 2021
Average flying height	~2531 m
Focal length	210.914 mm
Scan Resolution	15 cm (photo scale was 1:12,000)

Table 2. Airborne LiDAR sensor setting

LiDAR sensor setting	
Acquisition data	March 17th, 2005
Laser sensor	Optec ALTM 3100
Laser wavelength	1064 nm (Near-infrared red)
Laser point densisty range	3 to 20 points/m^2

Table 1. Camera settings of aerial photos

4. METHODOLOGY

4.1 Aerial triangulation process

Stereo photos were used to rectify the image to make orthophotos using a series of control and tie points. The aerial triangulation process depends upon known exterior orientation parameters determined by the relationship between image coordinates and known ground coordinates. These parameters determine the orientation of each aerial photo and the location of the perspective point (camera location). The aerial photos were triangulated using Leica Photogrammetric Suite (LPS, Leica Geosystems, Inc.).

4.2 LiDAR derived DSM

A 1 m by 1 m planimetric grid was generated over the research area. Within each grid cell, the local maximum heights of LiDAR points were distinguished and the maximum height value was stored at each grid location to create a Digital Surface Model (DSM). The DSM was smoothed by a 3x3 Gaussian filter (Hyyppä, et al., 2001) to reduce local variation on the surface. With this 1 m by 1 m smoothed DSM, the slope and aspect were calculated using the function of ArcGIS (ESRI Inc.). The slope and aspect were evaluated by the 3x3 neighboring pixels of DSM (Burrough and McDonnell, 1998).

4.3 Backprojection process

The ground coordinates of grid points from LiDAR derived DSM were transferred to photo-pixel coordinates using the collinearity equations below and an affine transformation. The collinearity equations are:

$$X_a = X_0 - f\left[\frac{m_{11}(X-X_{EX})+m_{12}(Y-Y_{EX})+m_{13}(Z-Z_{EX})}{m_{31}(X-X_{EX})+m_{32}(Y-Y_{EX})+m_{33}(Z-Z_{EX})}\right] \quad (1)$$

$$Y_a = Y_0 - f\left[\frac{m_{21}(X-X_{EX})+m_{22}(Y-Y_{EX})+m_{23}(Z-Z_{EX})}{m_{31}(X-X_{EX})+m_{32}(Y-Y_{EX})+m_{33}(Z-Z_{EX})}\right] \quad (2)$$

(Mikhail et al. 2001)

where m is the rotation matrix based on *omega (ω), phi (φ),* and *kappa* (κ) values of exterior parameters, the subscripts of m_{ij} are the ith row and jth column number of the rotation matrix m. X_a and Y_b are image space coordinates, X_0 and Y_0 are photo centers, f is the focal length, X_{EX}, Y_{EX}, *and* Z_{EX} are the coordinates of perspective centers, and X, Y, *and* Z are the sampled coordinates from LiDAR derived DSM.

To transfer from image space to photo pixel coordinates, an affine transformation (Mikhail et al. 2001) based upon the camera fiducial points was applied.

Note that the resolution of DSM (1 m) and the image pixel resolution of aerial photo (0.15 m) are different. To match the pixel resolution for this fusion technique, the different sizes of spatial filter were applied on the image to find the appropriate size. The following section describes more about the spatial filter.

4.4 Hillshade to cast the shadow on LiDAR derived DSM

With slope and aspect angle derived from the DSM, the shadow was cast using a hillshade function which is a modified version of the equation developed by Burrough and McDonnell (1998). This function is also used for ArcGIS hillshade function of spatial analysis.

$$\text{Hillshade} = 255 * ((\cos(\text{Zenith}) * \cos(\text{Slope})) + (\sin(\text{Zenith}) * \sin(\text{Slope}) * \cos(\text{Azimuth} - \text{Aspect}))) \quad (3)$$

(Burrough and McDonnell, 1998)

where Zenith is 90 – (altitude of the light source), Slope is the surface slope angle derived from the DSM, Azimuth is the azimuth angle of the light source, and Aspect is the surface

aspect derived from the DSM.

The hillshade surface was created in the view of sun and camera location. The hillshade surface used by ArcGIS uses the parallel light ray to calculate Zenith and Azimuth angle over the entire landscape, because the sun altitude is assumed to be at infinity and the azimuth angle is parallel from the sun. However, the camera location mounted on the airborne vehicle has a lower altitude and the azimuth is the radial direction from the camera location. Equation 4 was, therefore, customized to cast the shadow from the camera location in this study.

4.5 Image composition scheme using hillshade function

After respective true orthophotos were created from the three photos, the three true orthophotos were merged to make a composite image. In this study a novel image composition scheme was developed using visible side of hillsahde surface of each acquisition time. The camera view and the sun oriented hillshade surfaces were generated for all images. There were three sun oriented hillshade images and three camera view oriented hillshade images. Each hillshade pixel has the value ranged between 0 and 254. Generally, a hillshade value more than 177 implies the exposed/visible side of the surface towards the sun or camera view and a value less than 177 implies the obscured/occluded side of the surface. These hillshade values among images in the overlapped area were used to compare values and select the higher (or the highest) values to make a composite image from the respective orthophotos.

During the backprojection process, a spatial filter was applied to the filtered region centered at the exact location of drape points on the original perspective photo. This was done because the slight misalignment between draped location from the draped point and exact location should be considered to obtain the color from the original perspective photo. Four different filter sizes were used. The four were the exact location, a 3 by 3 pixels (3x3) filter, a 5 by 5 pixels (5x5) filter, and a 7 by 7 pixels (7x7) filter area (1 pixel cell size is 15 cm). These four cases were used for color acquisition from the original perspective photos.

5. RESULTS

5.1 Aerial Triangulation process of aerial photos

The accuracy from the block triangulation report is shown in Table 2. This shows the variations of GCPs and the location of tie points between each paired image. Table 2 shows the overall RMSE of the block triangulation as 37 cm. The error of comparison with the ground coordinates resulting from the triangulation approach were 37 cm, 48 cm, and 53 cm for the X axis, Y axis, and Z axis respectively.

Ground X coordinate error	37 cm
Ground Y coordinate error	48 cm
Ground Z coordinate error	52 cm
Image X coordinate error	0.13 mm
Image Y coordinate error	0.03 mm
The total RMSE	38 cm

Table 2. The accuracy assessment of aerial triangulation process.

5.2 True orthophoto generation

The improvement of this technique is shown with USGS orthophoto provided by the USGS National Map seamless server. Their orthophoto was captured in the same year (2005) as the aerial photos used in this study. The difference between USGS orthophoto and the true orthophoto was shown on Figure 3. To demonstrate the true projection, the rectified trees in the upright position on the true orthophoto, while trees in the USGS ortho photo were tilted only to show one side.

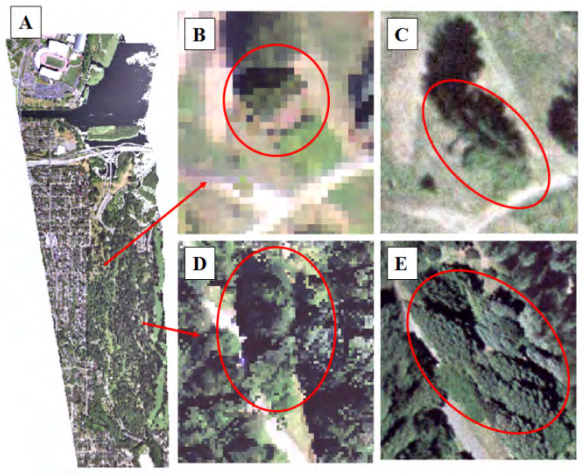


Figure 3. The true orthophoto generation. The final product of true orthophoto for the entire area (A). Single tree shown on the trueortho (B) and the same tree shown on USGS orthophoto (C). A group of trees shown on the trueortho (D) and the same trees shown on USGS orthophoto (E).

5.3 Filter size

Color was obtained for four cases: exact location of draped LiDAR points, a 3x3 filtered region centered at the location of draped LiDAR points on each photo, a 5x5 filtered region, and a 7x7 filtered region. The color variation is smaller and the noise of color values becomes smoother as the size of filter becomes larger. The 7x7 filter is, therefore, appropriate for this case because the color contrast is larger than for the other size of filters. The 7x7 pixel size filer (1 pixel size is 15 cm) is also close to the 1 m by 1 m cell size of LiDAR derived DSM used in this study. A composite image was made from three true orthophotos. Figure 3 shows the resulting composite image with the 7x7 filter and shows the composited true orthophoto for one tree and a group of trees.

6. DISCUSSION

6.1 Accuracy

The RMSE for all axes was less than 50 cm from the block triangulation report that is less error than that reported by St-Onge et al. (2004). The Z coordinates error was larger, because the surveying GPS unit is more error prone in the Z coordinate.

In this study, the conventional way to make an orthophoto was taken using field observed GPS for GCP during aerial triangulation process. GCP can, however, be identified and selected from the LiDAR three dimensional point cloud, if the point density is high enough to visualize objects. Liu et al. (2007) took this approach to derive GCPs from the aerial photos without the actual field survey. They found that LiDAR derived GCPs and DTM from LiDAR significantly reduced positional errors in the orthorectification process. If the GCP location is well identified by LiDAR itself, the error of the mismatch

between field measured GPS and LiDAR elevation acquired by INS can be decreased and the accuracy of the backprojection can be improved. With such an improvement, a smaller filter could be applied to obtain the colors from aerial photos during the composite process and the pixel resolution of DSM could be higher with high density LiDAR points.

With regard to the seasonal difference between the two datasets, the DSM was created with the leaf-off LiDAR data and aerial photos were captured in the leaf-on season. The seasonal difference may have caused some of the noise in the composite image. For example, the noise around trees along the road may be caused by this seasonal difference since the misalignment of the backprojection process was greater. It would be ideal to capture both the LiDAR data and aerial photos at the same time or in the same season.

6.2 Occlusion detection and compensation using hillshade function

As reported by Sheng et al (2003) and Kane et al. (2008), the LiDAR derived DSM does not represent the exact shape of the crown. Especially for parts lower than the lowest branches, the tree shape provided by the DSM is not realistic. To show this effect, Figure 4 shows the hillshade surface using the DSM.

The lower part of the crown had low hillshade values because the hillshade was calculated based on slope and aspect of the surface. If the DSM had abrupt slope change at the sides of tree crown shown on Figure 4, the slope and aspect were never exposed to the camera and sun direction. To show the effect, the edges of tree crowns had black frames which represent the lower values of hillshade in three dimensional view of Figure 4.

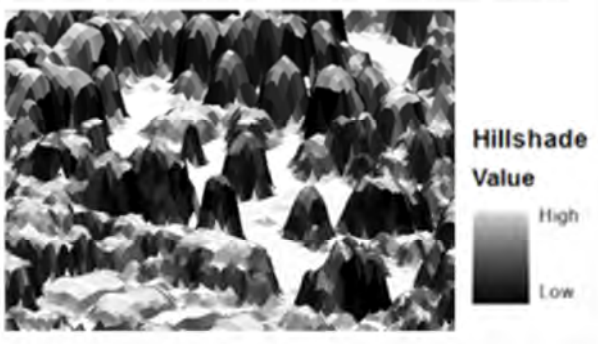


Figure 4. Shadowing effect using hillshade surfaces. Shadows are casted on the side of the DSM, clearly seen from the figure.

The hillshade value is calculated by slope and aspect angles which are evaluated by the gradient of the neighbouring pixels. The abrupt angles at the edge of tree canopy tend to be occluded area. If the tree canopy model is improved to cover the bottom of tree crown, more realistic hillshade values can be calculated and used for the composite process. The wrapped surface (Kato et al., 2009) can provide a more realistic surface to avoid this situation.

Hillshade surfaces provided useful information to find occlusions. If these hillshades are used for the orthorectification process of digital aerial photo with LiDAR taken at the same time, a fully automated true orthophoto process is possible.

6.3 True orthophoto

For the group of trees in Figure 3, the seam (cut) lines appeared on the USGS ortho photo. Conventionally, the final orthophoto tiles were mosaiced from more visually appealing parts of several individually rectified images. These mosaics will exhibit “seam lines” where the individual images met, and tree images at the seam boundaries did not match due to the displacement of tree tops. Even though the aerial photos have been controlled and rectified, the images of trees along the edges of tiles or at seam lines will appear tilted on orthoimages due to perspective angles and the use of DTMs that represent the ground elevations rather than the surface of the tree canopy. This resulted in an orthophoto where ground features were placed in the correct planimetric position, while the treetops were distorted horizontally. Most orthophotos are produced in this manner.

The data fusion using hillshade introduced here can be applied to the data fusion between LiDAR and the other optical spectral datasets at landscape level. The multi- and hyper-spectral images acquired by airborne sensors can have improved rectification with this true orthophoto creation process, even though they are not captured by stereo-pair. The high hillshade values of both camera view and sun angle can be used for spectral calibration. The advantage of this technique is that it selects the pure spectral signature of tree crowns for further species identification with less shadow and shaded pixels, because the exposed side of tree crowns is identified with camera view and sun oriented hillshade surfaces.

7. CONCLUSION

A technique of data fusion between aerial photos and LiDAR was developed. In particular, a true orthophoto was created using the LiDAR derived DSM and hillshade function on the DSM. To make the true orthophoto, backprojection process was used to transfer the ground coordinate to the photo coordinate on the perspective aerial photos using a collinearlity equation. However, to fuse LiDAR derived DSM with aerial photos, only the visible sides of DSM to photo perspectives were required to be identified and selected. The camera view and the sun oriented hillshade surfaces were generated using the LiDAR derived DSM and used to distinguish between the exposed and the occluded side of the DSM. An adapted technique can be applied to the data fusion between LiDAR and multispectral high resolution images, even though the images are not collected with full stereo coverage. If the camera or sensor location is reported, the hillshade function is useful to detect the occluded side of an object to compensate for and correct the color from aerial photos. The pure spectral signature collected with the hillshade surface can also be useful for species identification. Furthermore, the spectral value derived only from the visible side of DSM can be used for color calibration among aerial photos or multi-temporal spectral images. Therefore, this technique dramatically increases the opportunities for the data fusion between LiDAR and any spectral image for the image composition and color calibration.

REFERENCES

Amhar, F., Jansa, J., and Ries, C., Angel, E., 1998. The generation of true orthophotos using a 3D building model in conjunction with a conventional DTM. *International Archive of Photogrammetry and Remote Sensing*, 32 (Part 4): 16-22.

Angel, E., 2003. *Interactive Computer Graphics A Top-Down Approach with OpenGL* (pp. 559-593). NY: Addison Wesley

Brandtberg, T. and Walter F., 1998. Automated delineation of individual tree crowns in high spatial resolution aerial images by multiple-scale analysis. *Machine Vision and Applications* 11, 64-73

Burrough, P. A. and McDonell, R.A., 1998. *Principles of Geographical Information Systems*. New York NY: Oxford University Press

Erikson, M., 2003. Segmentation of individual tree crowns in colour aerial photographs using region growing supported by fuzzy rules. *Canadian Journal of Forest Research* 33, 1557-1563

Gougeon, F.A., 1995. A crown-following approach to the automatic delineation of individual tree crowns in high spatial resolution aerial images. *Canadian Journal of Remote Sensing* 21(3), 274-284

Habib, A.F., Kim E-M., and Kim C-J., 2007. New methodologies for true orthophoto generation. *Photogrammetric Engineering & Remote Sensing*, 73 (1): 25-36.

Hais, M. and Kučera, T., 2008. Surface temperature change of spruce forest as a result of bark beetle attack: remote sensing and GIS approach. *European Journal of Forest Research*, 127:327-336.

Hyyppä, J., Kelle, O., Lehikoinen, M., & Inkinen, M., 2001. A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. *IEEE Transactions on Geoscience and Remote Sensing*, 39(5), 969-975

Iverson, L.R., Dale, M.E., Scott, C.T., and Prasad, A., 1997. A GIS-derived integrated moisture index to predict forest composition and productivity of Ohio forests (U.S.A). *Landscape Ecology* 12: 331-348.

Kane, V. R., , Gillespie, A. R., McGaughey, R., James Lutz, J., Ceder, K., and Jerry F. Franklin, J. F., 2008. Interpretation and topographic correction of conifer forest canopy self-shadowing using spectral mixture analysis. *Remote Sensing of Environment* 112(10), 3820-3832.

Kato, A., Moskal, L.M., Schiess, P., Swanson, M.E., Calhoun, D., and Stuetzle, W., 2009. Capturing Tree Crown Formation through Implicit Surface Reconstruction using Airborne Lidar Data, *Remote Sensing of Environment* 113: pp1148-1162

Leckie, D., Gougeon, F., Hill, D., Quinn, E.,Armstrong, L., and Shreenan, R., 2003. Combined high-density lidar and multispectral imagery for individual tree crown analysis. *Canadian Journal of Remote Sensing* 29(5), 633-649

Leckie, D.G., Gougeon, F.A., Tinis, S., Nelson, T., Burnett, C. N., and Paradine, D., 2005. Automated tree recognition in old growth conifer stands with high resolution digital imagery. *Remote Sensing of Environment* 94, 311-326

Liu, X., Zhang, Z., Peterson, J., and Chandra, S., 2007. LiDAR-derived high quality ground control information and DEM for Image Orthorectification. *Geoinformatica* 11, 37-53

Mahiny, A.S. and Gholamalifard, M., 2007. Dynamic spatial modeling of urban growth through cellular automata in a GIS environment. *International Journal of Environment Research* 1(3): 272-279.

Mikhail, E.M., Bethel, J.S., and McGlone, J.C., 2001. *Introduction to modern photogrammetry* New York, NY: John Wiley & Sons, Inc.

Pellegrini, L., Boni, P., and Carton, A., 2003. Hydrographic evolution in relation to neotectonics aided by data processing and assessment: some examples from the Northen Apennines (Italy). *Quaternary International* 101-102: 211-217.

Rau, J-Y., Chen, N-Y, and Chen L-C., 2002. True orthophoto generation of built-up areas using multi-view images. *Photogrammetric Engineering and Remote Sensing* 68 (6): 581-588.

Rönnholm, P., Hyyppä, J.,Hyyppä, H.,Haggrén, H., Yu, X., and Kaartinen, H., 2004. Calibration of laser-derived tree height estimates by means of photogrammetric techniques. *Scandinavian Journal of Forest Research* 19, 524-528

Schickler W. and Thorpe A., 1998. Operational procedure for automatic true orthophoto generation. *International Archive of Photogrammetry and Remote Sensing* 32 (Part 4): 527-532.

Sheng Y., Gong, P. and Biging, G.S., 2003. True orthoimage production for forested areas from large-scale aerial photographs. *Photogrammetric Engineering and Remote Sensing* 69 (3): 259-266.

St-Onge, B., Jumelet, J., Cobello, M., and Véga, C., 2004. Measuring individual tree height using a combination of stereophotogrammetry and lidar. *Canadian Journal of Remote Sensing* 34, 2122-2130

Van Den Eeckhaut, M, Poesen, J., Verstraeten, G., Vaanacker, V., Moeyersons, J., Nyssen, J., and van Beek, L.P.H., 2005. The effectiveness of hilldshade maps and expert knowledge in mapping old deep-seated landslides. *Geomorphology* 67: 351-363.

Van Den Eeckhaut, M., Poesen, J., Verstraeten, G. Vaanacker, V., Nyssen, J., Moeyersons, J., van Beek, L.P.H, and Vandekerckhove, L., 2007. Use of LIDAR-derived images for mapping old slides under forest, *Earth Surface Processes and Landforms* 32: 754-769.

Wulder, M., Niemann, K.O., and Goodenough, D.G., 2000. Local maximum filtering for the extraction of tree locations and basal area from high spatial resolution imagery. *Remote Sensing of Environment* 73, 103-114

Zhou, G., Chen, W., Kelmelis, J.A., and Zhang, D., 2005. A comprehensive study on urban true orthorectification. *IEEE Transactions on Geoscience and Remote Sensing* 43 (9): 2138-2147.

ACKNOWLEDGEMENTS

The authors would like to express our sincere gratitude to Dr. Ward W. Carson of the former USDA Forest Services and American Society of Photogrammetry and Remote Sensing for giving two awards (2008 Ta Liang Award and 2008 Paul R. Wolf Scholarship) to support this research.

MAPPING OF BUILT–UP AREA DENSITY FROM SATELLITE IMAGES USING MORPHOLOGICAL GRANULOMETRIES

A. Kemmouche [a]*, R. Khedam [a], C. Mering[b]

[a] Laboratoire de Traitement d'Images et de Rayonnement, Faculté d'Electronique et d'Informatique, Université des Sciences et de la Technologie Houari Boumediene, BP 32, Bab-Ezzouar 16111 Alger, Algérie, akemmouche@hotmail.com

[b] Laboratoire Pôle Image, UFR GHSS, case courrier 7001, Université Paris 7-Denis Diderot, 2 Place Jussieu, 75251 Paris Cedex 05 France, mering@lgs.jussieu.fr

KEY WORDS: Mapping, Urban, Mathematics, Analysis, Classification, Image

ABSTRACT:

The overall objective of this work is to provide maps based on the spatial organization of built-up areas and to achieve the comparative spatial analysis of built-up areas on east of Algiers in 1985 and 1996. Landsat TM images from both dates are processed here in order to characterize spatial and temporal change in built-up areas. Contextual supervised classification method is used for built-up areas extraction. Built-up density mapping is provided by local granulometric analysis, based on binary mathematical morphology. This method enables the classification of entities according to their granulometric descriptors generated by opening granulometries.

1. INTRODUCTION

Built-up areas in Algiers have markedly increasing during the last decades. Growth of urban built-up area is accompanied by an evolution of land use. Remote sensing images are relevant materials for observation and thematic mapping by multispectral and multi-textural classification. The objective here consists in mapping the spatial organisation of one single component of the landscape under study, such as built-up areas. Many techniques have been developed for built-up analysis (Zhang, 2002; Zha, 2003). Different parameters can be used to define the spatial organisation of a set: the size of convex entities forming the set, their form or their ordering. Texture is the characterizing feature of built-up areas in satellite imagery. For some researchers (Matsuyama, 1983; Wood-1996), there exist two categories of methodologies to analyze the textures, a statistical and a structural one. The former model the textures as a random function without a regular structure and are utilized for the detailed textures. The structural methods describe the textures produced by the regular structure of textural elements (Philipp, 1994).

In the local or global analysis of the texture, the textural parameters computed from the local or global statistics of the image with grey tone images are used as classifying descriptors. However, computation of the texture parameters from grey-tone values is not relevant for the feature extracted data used in this study. The work presented in this paper is about the development of a methodology for the quantification of the spatial organization of built-up areas from binary images. Such spatial organization is called macro-texture. On a binary image, Busch *et al*. (Busch, 1998) define the feature density as the number of pixels matching this feature that are contained in an image window. The proposed approach is based on mathematical morphology (Serra, 1982; Soille, 2003), and has been used successfully in vegetation density mapping (Kemmouche, 2004). In a first step, after extraction of the built-up areas by multi-spectral analysis, we define the descriptors of the macro-texture from the concept of the granulometric analysis of binary images representing built-up areas. To describe the macro-texture around a pixel of the image, we have calculated the local granulometric density over a window centred on the pixel. The result of this granulometric computation is a vector associated with each pixel and new grey level images generated.

In a second step, all the pixels of the original, described by the macro-texture parameters are classified by the K-means method to produce the final map, which can be considered as a map of the density of the built-up areas. The method has been applied to map built-up areas density from satellite data on east of Algiers in 1985 and 1996. Such maps are efficient tools to study the spatial dynamics of built-up areas.

The paper is organized as follows. In section II, the classification of satellite data with high spatial resolution from urban areas is described. The proposed mathematical morphology approach to built-up areas density mapping is discussed in section III. Experimental results are given in section IV.

2. BUILT-UP AREAS EXTRACTION

In this section image analysis methods used for extraction of built-up areas from Landsat images are described. The first part is devoted to image processing adapted to urban areas classification from multispectral images, and the second part describes built-up areas extraction. Landsat images corresponding to seven band multispectral mode (Thematic Mapper) were explored over eleven-year period.

2.1 Classification of urban areas from satellite data

There are many different approaches to classifying remotely sensed data. They all fall under two main topics: unsupervised *and* supervised classification. Supervised classification methods are two kinds: punctual or blind methods and contextual methods (Pieczynski, 1989; Richards 1993). Punctual classification methods are conventional classification

techniques which classify each pixel independently by considering only its observed intensity vector. The result of each method has often a "salt and pepper" appearance characterizing misclassification. It means that intensity vector is insufficient and then leads to incorrect classification of pixels. In particular of remotely sensed data, adjacent pixels are related or correlated, both because imaging sensors acquire significant portions of energy from adjacent and because ground cover types generally occur over a region that is large compared with the size of a pixel. Using coherent contextual information for classification efficiency and accuracy in remote sensing has long been desired. Contextual information is important for the interpretation of a scene. When a pixel is considered in isolation, it may provide incomplete information about the desired characteristics. However, the consideration of the pixel in its context, more complete information might be derived. The basic idea of spatial context is that the response and class of two spatially adjacent pixels are highly related. For example, if *(i, j)* and *(m, n)* are two neighbouring pixels and if *(i, j)* belongs to class *k*, then there is a high possibility that pixel *(m, n)* also belongs to the same class *k*. Therefore, the decision for a pixel is taken based not only on the observation at *(i, j)* but also on all observations at *(m, n)* where *(m, n)* is neighbour of *(i, j)*. Among contextual methods, the most widely applied to remote sensing images is the Markov random Field (MRF) approach , which has given very promising results (Schistad 1999a; Schitad 1996b, Khedam 2001). MRF is given as the best methodological framework to describe the correlation of neighbouring pixels.

2.2 MRF contextual classification model

We assume that a classified image *X* and observed data *Y* are realisations of stochastic processes ***X*** and ***Y***, respectively. $Y = \{Y^1, Y^2, ..., Y^K\}$ are multispectral data observed through *K* spectral bands and are supposed to be acquired on a finite rectangular lattice $W = \{s = (i, j) : 1 \le s \le S\}$, *s* is the site of the *ijth* pixel and *S* is lattice's area. The multispectral data can be described with $Y = \{y_s, 1 \le s \le S\}$ where $y_s = \{y_s^1, y_s^2, ..., y_s^K\}$ is a feature vector observed on the site *s*. Our goal is to find the optimal classified image $X^* = \{x_s, ..., x_S\}$ based on the observed data *Y*. Each site of the segmented image is to assigned into one of *M* classes; that is, $x_s = \{1, 2, ..., M\}$ where *M* is the number of classes assumed to be known in supervised classification process. This optimisation is executed from the view point of the maximum *a posterior* (MAP) estimation as follows:

$$X^* = X_{MAP} = \underset{X \in \Omega}{argmax}\{P(X/Y)\} \quad (1)$$

Where Ω is labelled configurations set. Following Bayes theorem, equation (1) becomes:

$$X_{MAP} = \underset{X \in \Omega}{argmax}\left\{\frac{P(Y/X)P(X)}{P(Y)}\right\} \quad (2)$$

The modelling of both class conditional distribution *P(Y/X)* and prior distribution *P(X)* becomes an essential task. *P(Y)* is the probability distribution of the observed data and doesn't depend on the labelling *X*. Note that the estimate (2) becomes the pixel–wise non–contextual classifier if the prior probability doesn't have any consequence in formulating (2). *P(Y/X)* is the conditional probability distribution of the observation *Y* given the labelling *X*. A commonly used model for *P(Y/X)* is that the feature vector observed Y_s is drawn from a *"Gaussian distribution"*. For a Markov random field *X* and so, according to the Hammerslay-Clifford theory, *P(X)* can be expressed as a Gibbs distribution with *"Potts model"* as energy function model. The global MAP estimate given by equation (1) is equivalent to the minimisation of the followed *a posterior* global energy function:

$$X_{MAP} = \underset{X \in \Omega}{\arg\min}\left\{U\left(X/Y\right)\right\} \quad (3)$$

Once MAP classification problem is formulated as an energy minimisation problem, it can be solved by an optimisation algorithm. Among the most effective algorithms for optimisation in the framework of image MRF modelling are Simulated Annealing (SA) (Geman, 1984) whose the computational demands are well known and Iterated Conditional Modes (ICM) (Besag, 1986) which is a computationally feasible alternative of the SA with a local minimum convergence of the energy function. To use ICM algorithm, global minimisation energy function of equation (3) must be transformed on the followed local minimisation energy function:

$$U(x_s/y_s) = \left\{\left\{\frac{1}{2}(y_s - \mu_{x_s})^T . \Sigma_{x_s}^{-1} . (y_s - \mu_{x_s}) + \frac{1}{2} ln\left\{\left|\Sigma_{xs}\right|\right\}\right\} + \beta \sum_{r \in V_S} (1 - \delta(x_s, x_r))\right\} \quad (4)$$

Where μ_{xs} and Σ_{x_s} are class *xs* are respectively mean vector and covariance matrix of class x_s estimated during training process. β is a regularisation parameter and is frequently user specified. δ is Kroeneker symbol calculated on the neighbourhood *Vs* of site *s*.

ICM algorithm can be resumed on five steps as follows:

Step 1: Estimate statistic parameters set (μ_{xs}, Σ_{x_s}) from the training samples of each class from M classes

Step 2: Based on μ_{xs} and Σ_{x_s}, estimate an initial classification using the non-contextual pixel-wise maximum likelihood decision rule. We use the first term of equation (4)

Step 3: Choose an appropriate value of β, an appropriate shape and size of neighbourhood system *Vs* and an appropriate convergence criterion.

Step 4: Perform the local minimisation defined by equation (4) at each pixel in specified order: update *ys* by the class *xs* that minimises equation (4)

Step 5: Repeat step (3) until convergence.

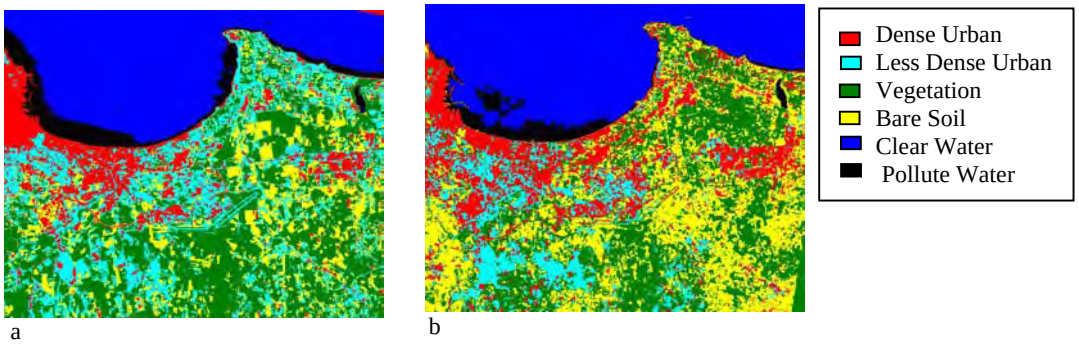
2.3 Built-up area extraction from classified urban areas

The described algorithm is applied to classify satellite images of the selected region of interest. Multispectral and multitemporal images were acquired in 1985 and 1996 by

ETM+ sensor of Landsat-7 satellite. The images cover the north-eastern part of Algiers (Algeria). The RGB compositions of these two images are given on figure 1. Six thematic classes dominate the study site: Dense Urban (DU), Bare Soil (BS), Less Dense Urban (LDU), Vegetation (V), Clear water (CW), and Pollute Water (PW). Using a 2-D scatterogram of ENVI software, data samples are selected automatically from each class for training and testing the proposed classifier. The MRF contextual classification results (8-connexity and $\beta = 0.75$) are shown on figure 2. Statistical assessment of these results relatively to the considered test data gives an appreciate KHAT parameter of 91.6% for data acquired on 1985 and 90.2% for data acquired on 1996.

a b

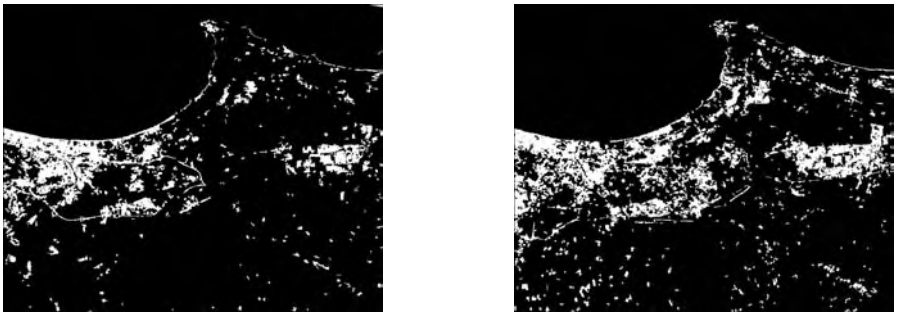
Figure 1. RGB composition of ETM+ images for 1985(a) and to 1996(b) scenes



a b

Figure 2. MRF classification result for 1985(a) and to 1996(b) scenes

From the obtained classified images (Figure 2), built-up area is extracted using a simple masking operation. Except dense urban (DU) class, all the other classes (BS, LDU, V, CW, PW) are masked which means that except DU pixels, all other pixels are assigned label “0”. The resulting binary images are presented in figure 3 (a and b) for both dates 1985 and 1996.



a b

Figure 3. Built-up areas extracted from TM scene corresponding 1985(a) and to 1996(b)

3. GRANULOMETRIC ANALYSIS FOR QUANTIFICATION OF THE BUILT-UP DENSITY

The process of built-up areas density mapping is organised in two parts. In the first one, granulometric analysis is computed on binary images with built-up areas in order to define the macro-texture parameters. In the other part, a density map is built by automatic classification of granulometric images.

3.1 Granulometric analysis on binary images

A binary image can be described as a set of the Euclidean space R^2. Such a set consists of many subsets, which are the connected components of the image. We choose here a criterion, which is the size distribution of the subsets in order to perform the textural analysis of the set. It is obtained by global transformations and measurments on the image. This analysis called Granulometric analysis is very similar to quantitative analysis of soil granulometry by sieving and weighting. The concept of granulometry for Euclidean set analysis was introduced by Matheron (Matheron, 1967) as a new tool for studying porous media. The principle of binary morphological granulometric size distributions was conceived by Matheron (Matheron, 1975) as a way to describing image granularity. The sieving of grains within the image was accomplished by a series of morphological openings with convex structuring element of increasing size.

A series of openings $O^{\lambda B}$ with a family of structuring elements B_1, B_2, ... B_n, is a granulometry if it satisfies the following axiom:

$$\forall (i, j)\ i \le j \Rightarrow O^{B_i} \ge O^{B_j} \quad (6)$$

Many parameters are provided by granulometric analysis, such as *granulometric distribution.* In order to assess the size distribution of the connected components of a set X, we use the method of granulometry by opening with a convex structuring element B. It consists in a successive application of morphological openings on the set X using an increasing structuring element B. As the size of the structuring element B increases, more and more details in the image are suppressed. The connected components, which are smaller than B, are eliminated. By increasing the size λ of B, the elements of size $(\lambda - 1)$ are successively eliminated as though they were sieved. Computation of the area of elements suppressed at each opening step on the whole image leads to the evaluation of the size distribution $G_\lambda(X)$ of the set X, which is:

$$G_\lambda(X) = \left[A(X) - A\left(O^{\lambda B}(X)\right)\right] / A(X) \quad (7)$$

where, $A(X)$ indicates area on the initial image and $A\left(O^{\lambda B}(x)\right)$ is the area of the set X opened by structuring element of size λ.

Experiments such as studies on porosity of rocks from thin section images (Serra 1982) show that, in case of finite sequence of openings (i.e. for finite values of λ), the granulometric density is more relevant than the granulometric distribution for providing-g efficient descriptors of the size of the components of the binary image. The granulometric density $g_\lambda(X)$ of a binary image $X \subset R^2$ relative to a convex structuring element B is defined as:

$$g_\lambda(X) = \left[A\left(O^{\lambda B}(X)\right) - A\left(O^{(\lambda+1)B}(X)\right)\right] / A(X) \quad (8)$$

Granulometric density $g_\lambda(X)$ represents the fraction of total area of X that is rejected between two successive openings of respective radius λ and $(\lambda + 1)$. It provides a statistical evaluation of the area of the components of X : the maxima of $g_\lambda(X)$ indicate that there are a high proportion of subsets of X having a radius inferior to λ.

This analysis generates a finite and homogeneous set of quantitative descriptors $g_\lambda(X)$ that can be easily used to quantify the density.

3.2 Computation of macro-textural descriptors

The application of granulometric analysis has focused on taking local granulometric density around individual pixels. The use of local granulometric analysis to define texture descriptors was introduced by Dougherty et al (Dougherty, 1992; Chen, 1992). Rather than computing granulometric density across an entire image, as the global granulometry, pixel counts are only taken locally in windows about each pixel, thereby generating local granulometric density at each pixel: for each pixel of the binary image the *local granulometric density* $g_\lambda(X)$ was measured over a window centred on the pixel P at each stage of the granulometry**.**

This value is computed for the windows $F(P)$ around all the pixels P. The resulting texture representation is a vector $\{V_i(P)\}$. The texture variables describing a pixel P will be $\{V_i(P)\}_{i=1,2,.....,I}$ where,

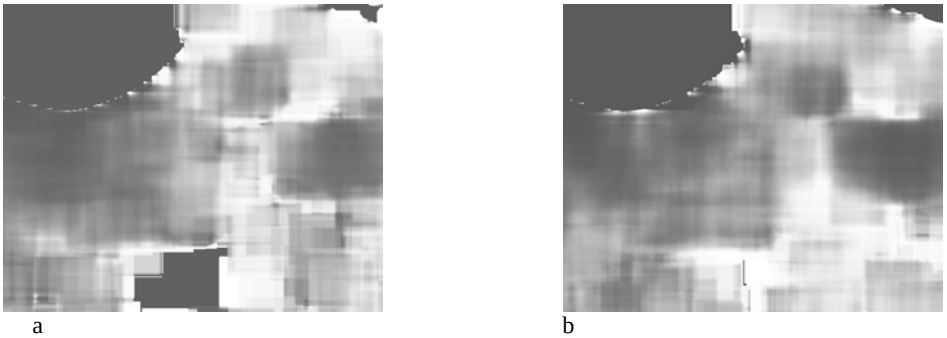
$$V_i(P) = g_i(F(P)) \quad (9)$$

$g_i(F(P))$ is the value of the local granulometric density computed inside the window $F(P)$ centred on pixel P at opening step i.

I is the size of B such that all the pixels of $F(P)$ are eliminated after the opening by B_I.

By this way, each pixel P of the binary image is described by the I values of Vi where, $i = 1,2,....,I$. Local granulometric densities for all the pixels of the binary image were computed and results were reassembled to form an image. Such a processing is performed for $i = 1,2,....,I$ then I gray-tone images are generated. This set of images will then be used as input for the classification step.

Seven images of granulometric density were obtained by this way. An example is shown on figure 4, which corresponds to the image of the granulometric density of size $\lambda = 1$ computed from the binary images of built-up areas of figure 3.



a b

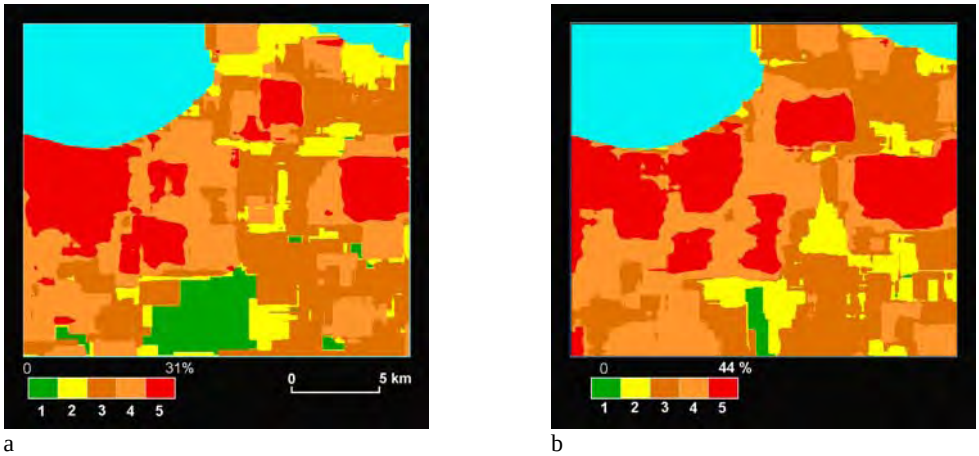
Figure 4. granulometric density images computed from binary images of built-up in figure 3

3.3 Built-up density mapping from macro-textural indicators

In order to obtain a built-up density map, we applied granulometric image processing and produced I gray tone images from binary images of built-up areas extracted on section 2.3. The map is then obtained by a multi-channel classification on density ganulometric images. An unsupervised classification of each pixel is performed by a K-means method (Diday, 1974). Classification of macro-texture at a pixel P is based upon the descriptor vector of granulometric densities at P. The 'macro-textural descriptors' are the input variables for the classification process. The result is a k-colours image. Each class is interpreted according to the mean granulometric density values and it contains pixels having similar macro-textural signatures. When the neighbouring contains only small components, it corresponds to high values for smaller size of sieving. At the opposite, it may correspond to high value for biggest size of sieving, if the neighbouring contains mainly large components. Such an analysis leads to the legend of the map in terms of density. The classes are coloured with a red to green colour scale to show the progressive decreasing in the density of built-up zones. For both dates green colour represents the smaller built-up areas while red colour corresponds to the bigger areas.

4. RESULTS

We have analysed the macro-texture of the two binary images (figure 3 a and b) and mapped the different types of macro-texture. K-means classification was performed for both dates into five classes of built-up density. The result of these maps according to the density is represented in figure 5 (a and b).



a b

Figure 5- Built-up density map computed from binary images in figure 3 for 1985 (a) and 1996 (b)

The five classes show that progressive decreasing density can be summarized as follow:

Class 1: this class corresponds to nearly bare soils.

Class 2 : this corresponds to sectors of transition between areas essentially of bare soil to those with weak density.

Classes 3, 4 and 5: representing areas from intermediate to highest built-up density.

To differentiate the classes which represent the different densities of urban zones in Algiers, it is possible to use the notion of covered area over each of the two classes, the mostly dense and the least dense, for both dates. The area of the least dense class covers only 7.68% of the territory for the year 1996, while in 1985 it covers 11.5%. The occupation of the dense class represents only 13% in 1985 compared to areas occupying

16% in 1996 (see Table 1). The result of these analyses shows that the method of quantification of the density of the urban space using satellite images makes it possible to separate urban zones based on their density.

Class	Very dense	Least or not dense	
Number of structures	17	44	Algiers 1985
Number of pixels	75034	64567	
Percentage	13 %	11.5 %	
Number of structures	11	29	Algiers 1996
Number of pixels	93017	43033	
Percentage	16 %	7.68 %	

Table 1 Comparative result of the highest and lowest dense classes

5. CONCLUSION

The above-described method maps the built-up areas organizations by using the macro-textural descriptors of the classified images. In the produced map, the patterns are characterized by their relative macro-textures. The method has been applied for analysing and mapping the spatial variations of built-up areas using Landsat TM multispectral data. It provides a striking illustration of spatial organisation of urban zones from binary images. The produced maps for two dates leads to the analysis showing the evolution of built-up density during the period under study. The works will be focused in reproducing this method at a regional scale, in order to study built-up growth on the Algerian littoral by satellite image processing.

References

Besag, J., 1986. On the Statistical Analysis of Dirty Pictures. *Journal Royal of Statistics: Soc. B*, 48, 3, pp. 259-302.

Busch A., 1998. Revision of built-up areas in a GIS using satellite imagery and GIS data. In: D. Fritsh, M. English & M. Sester, eds, IAPRS, vol. 32/4, ISPRS Commission IV Symposium on GIS – Between Visions and Applications, Stuttgart, Germany.

Chen Y., Doughety E. R., 1992. Texture classification by gray-scale morphological granulometries. *Visual Communications and Image Processing*, SPIE vol. 1818.

Diday E., 1971. La méthode des nuées Dynamiques. *Rev. Stat. Apll.*, vol.19, n°2, pp. 19-34.

Dougherty, E. R., Pelz, J. B., Sand, F., Lent, A., 1992. Morphological image segmentation by local granulometries size distributions. *Journal of Electronic Imaging*, 1(1).

Geman, S., Geman, D., 1984. Stochastic Relaxation, Gibbs Distributions, and the Bayesain Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6, 6, pp. 721-741.

Kemmouche A. Mering C., Sansal B. and Dewolf Y., 2004. Macro-texture Mapping from satellites images by morphological granulometries: Application to vegetation density mapping in arid and semi-arid regions. *International Journal of Remote Sensing*, 25(23), pp. 5319-5335.

Khedam, R., Belhadj-Aissa, A., 2001. General Multisource Contextual Classification Model of Remotely Sensed Imagery based on MRF. *IEEE / ISPRS Workshop on Remote Sensing and Data Fusion Over Urban Areas, Rome, Italy, November 8-9th 2001*

Matheron, G., 1967. *Eléments pour une Théorie des milieux poreux*. Masson, Paris.

Matheron, G., 1975. *Random Sets and Integral Geometry*. Wiley, New York.

Matsuyama T., Miura S., Nagao M., 1983. Structural Analysis of natural textures by Fourier Transformation, *CVGIP*, 24, pp. 347-362.

Philipp S., Smadja M., 1994. Detection of surface specific points by local parallel processing of discrete terrain elevation data, CVGIP, vol.4, pp. 375-387.

Pieczynski W., 1989. Estiamtion of context in random fields, *Journal of Applied Statistics*, vol. 16, no. 2, pp. 283-289.

Richards J., A., 1986. *Remote sensing digital image analysis.* Springer Verlag, Berlin, 281p.

Schitad Solberg A. H., 1999. Contextual Data Fusion Applied to Forest Map Revision. *IEEE, Trans. Geos. Remote Sensing*, vol. 37, no. 3, pp. 1234-1243.

Schistad Solberg, A., H., Taxt, T., A., Jain, K., 1996. A Markov Random Field Model for Classification of Multisource Satellite Imagery. *IEEE, Trans. Geos. Remote Sensing*, vol. 34, no. 1, pp. 100-112.

Serra, J., 1982. *Image Analysis and Mathematical Morphology*. Academic Press, London, 610p.

Soille P., 2003. *Morphological image analysis: principles and applications*. 2nd ed. Springer-Verlag, Berlin, Germany.

Wood, 1996. *The geomorphological characterisation of Digital Elevation Model*, Thèse de doctorat, Université de Leicester (UK).

Zha Y., Gao J., Ni S., 2003. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *Int J Remote Sensing*, vol. 24, no. 3, pp. 583-594.

Zhang Q., Wang J., Peng X., Gong P., Shi P., 2002. Urban built-up land change detection with road density and spectral information from multi-temporal Landsat TM data. *Int J Remote Sensing*, vol. 23, no. 15, pp. 3057-3078.

ANALYSIS OF BRDF CHARACTERISTICS OF FOREST STANDS WITH A DIGITAL AERIAL FRAME CAMERA

T. Koukal [a, *], W. Schneider [a]

[a] Institute of Surveying, Remote Sensing and Land Information, University of Natural Resources and Applied Life Sciences, Vienna, Austria
(tatjana.koukal, werner.schneider@boku.ac.at)

KEY WORDS: Digital, Aerial, Photography, Land Cover, Forestry, Radiometry, Analysis, Modelling

ABSTRACT:

The directional reflectance characteristics of the land surface can be described based on concepts of the Bidirectional Reflectance Distribution Function (BRDF). This contribution concentrates on directional reflectance effects in aerial photos of forests. Since the advent of digital metric aerial photography, major efforts have been made to apply quantitative digital automatic methods for the analysis of aerial photos. Directional reflectance effects are important for this: on the one hand, they make analysis more complicated, on the other hand they may provide additional information for deducing land cover (e.g. forest) parameters. Digital aerial photos, when taken with high forward and side overlap, may provide a convenient tool for analysing directional reflectance effects. The main aim of this contribution is to demonstrate the usefulness of digital aerial photos taken with a Vexcel UltraCamD for analysing directional reflectance characteristics of forests. 11 BRDF models are tested for 6 different land cover types focusing on forest cover. The models were evaluated using the coefficient of determination (R^2) and the symmetric mean absolute percentage error (sMAPE). It has been shown that the parameters of BRDF models describing individual forest plots can be estimated from digital aerial photos taken with a frame camera with large forward and side overlap. Differences in the performance of the models for different forest plots could be explained taking into account the special assumptions on which the models are based and the special properties of the observed forest stands.

1. INTRODUCTION

In the analysis of remotely sensed images, information on land cover is mainly deduced from the reflectance properties of the terrain surface. Reflectance is a function of wavelength and of irradiation and observation directions. These directional reflectance characteristics can be described based on concepts of the Bidirectional Reflectance Distribution Function (BRDF) (Schaepman-Strub et al., 2006).

While BRDF aspects often can be neglected in the analysis of satellite images sensed with vertical view direction and with a small angular field of view (e.g. of Landsat images), the directional reflectance properties are important for image data from sensors with steerable view direction (e.g. Ikonos) and from sensors with a large angular field of view (e.g. MODIS, or aerial scanners and aerial cameras). BRDF effects are particularly noticeable in the case of surfaces of a pronounced vertical structure, e.g. forest canopies.

This contribution concentrates on the directional reflectance effects in aerial photos of forests. A quantitative treatment of this subject is of interest for the following reasons:

(1) In the past, aerial photos were recorded in analogue form on photographic film. They were of poor radiometric quality. Digital automatic image analysis therefore did not seem very promising and was hardly applied operationally for the evaluation of aerial photos. Rather, the usage of aerial photos (e.g. in forestry) primarily relied on labour-intensive and subjective visual interpretation. It is only now, after the advent of digital metric aerial photography, that major efforts are made to apply quantitative digital automatic methods also for the analysis of aerial photos. The BRDF-related effects (including the well-known hot spot phenomenon) in aerial photos which are usually taken with a large angular field of view pose a major challenge in this undertaking. These effects on the one hand make analysis more complicated, on the other hand they may provide additional information for deducing land cover (e.g. forest) parameters. It is desirable to expand the knowledge in this field.

(2) Directional reflectance data on different types of land-cover usually are obtained by measurements either in the laboratory on small samples or in the field by employing special platforms and constructions. It is impossible to perform directional measurements on forest canopies in the lab, and it is very difficult to implement outdoor directional reflectance measurements on timber trees and old stands. However, digital aerial photos, when taken with high forward and side overlap, provide a convenient tool for analysing directional reflectance effects. It is an additional special advantage of using aerial photography that analysis may be performed at different scales: Working with full resolution of the aerial photos (e.g. pixel sizes of 20 cm to 100 cm), one may study directional reflectance at individual trees and parts of trees (sunlit branches, treetops, etc.). When using reduced resolution (by averaging over pixel windows of the order of 20 m x 20 m or 50 m x 50 m), directional reflectance characteristics including the shadow effects of forest canopies due to crown shape, crown closure and ground vegetation may be analysed. On the other hand, a limitation of aerial photography is given by the restriction to (usually) 4 spectral bands in the visible and near infrared parts of the spectrum.

The main aim of this contribution is to demonstrate the usefulness of digital aerial photography with frame cameras for analysing BRDF-related characteristics of forests. In particular, a number of BRDF models are tested for their usefulness to describe the directional patterns of pixel values as obtained with a Vexcel UltraCam aerial frame camera. The directional reflectance functions obtained can be considered as

* Corresponding author

approximations to BRDF only, as the hemispherical irradiance component (diffuse sky radiation) is present in addition to the directional irradiation from the sun.
Large windows of pixels are used, causing the directional reflectance functions to include shadow effects due to crown shape, crown density etc.

2. METHOD

Basically there are two possibilities to employ wide-angular-view-sensors for studying directional reflectance effects: Single images may be used, if large homogeneous stands covering the entire area of an image are available. In this case, a sample of plots distributed over one image (and therefore viewed from different directions) is analysed, and the BRDF model parameters are deduced from the pixel values of this sample (Dymond and Trotter, 1997). Alternatively, if images with sufficient forward and side overlap are available, the same plot on the ground is viewed from different directions on different images. In this second case, the BRDF model parameters of every single terrain element may be derived. No assumptions on the homogeneity of stands beyond the area of an individual plot are required. This second approach was chosen for the work described here.
There are two different types of digital camera systems available: line cameras (e.g. ADS40 by Leica Geosystems) and frame cameras (e.g. DMC by Intergraph/ZI-Imaging, UltraCamX by Vexcel Imaging). In BRDF analysis, the number of different directions of view on a terrain element that can be obtained at one overflight is decisive. In case of a line scanner, this number is determined by the number of CCD lines in the focal plane. E.g. the ADS40 camera is a 3-line-camera leading to 3 different directions of view per terrain element and overflight. In case of a frame camera, the number of directions of view is dependent on the percentage of forward overlap. Digital frame cameras achieve forward overlaps of 90 % and higher and, as a consequence, 10 directions of view and more.
The number of viewing directions also increases with the overlap between the flight lines (side overlap). E.g. in a mission with 90 % forward overlap and 30 % side overlap, any terrain element is observed from 10 to 20 view directions depending on the object's position within the path (Figure 1).

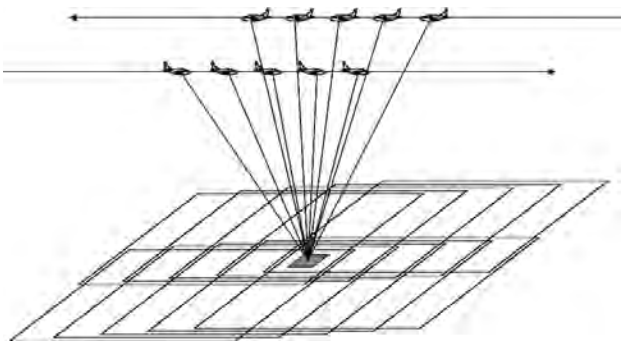


Figure 1. Directions of view on a terrain element in aerial photography with forward and side overlap

Another key parameter for BRDF-related investigations is the angular field of view (FOV) both along and across the track as it defines the range of viewing angles (from vertical) at which the terrain elements can be observed.

In this study, images taken with the digital metric camera Vexcel UltraCamD were used. Relevant parameters of camera and images are listed in chapter 4.
While the input data (pixel values) for many BRDF models may be given in arbitrary units, models accounting for multiple scattering require reflectance values for input (e.g. the WAK model by Dymond et al. (2001)). For this reason, the image data were (in an approximation) radiometrically calibrated, i.e. converted to reflectance values.
The empirical line method was used (Smith and Milton, 1999). Reflectance values of reference plots were measured on the days of the image flights. Reference plots were selected for which Lambertian reflectance characteristics can be assumed, e.g. bright surfaces of concrete roads and dark surfaces of asphalt roads. The radiance values measured at these surfaces were converted to reflectance values by comparison (ratioing) with radiance values measured at a horizontal Spectralon reference panel of known reflectance. As the images used for this study were taken at different times, the differences in sun zenith angle at these times had to be taken into account. The following algorithm was used for this:

$$p_{iO} = p_i \frac{\cos\theta_{SO}}{\cos\theta_{Si}} = a + b\rho_i \qquad (1)$$

$$\rho = \frac{p\frac{\cos\theta_{SO}}{\cos\theta_S} - a}{b}$$

Here, p_i are the pixel values from the positions of the plots on the images and ρ_i are the corresponding reflectance values from the terrestrial spectroradiometric measurements. θ_{Si} is the zenith angle of the direction to the sun at the time the aerial photo showing the pixel value p_i was taken, and θ_{SO} is a standard value of the zenith angle of the sun used as a common reference: All pixel values of the plots for radiometric calibration, p_i, are converted to fictitious values p_{iO} they would have for this standard sun zenith angle. a and b are the coefficients of linear regression of the values p_{iO} to the values ρ_i. With this model, any pixel value p from an image taken at sun zenith angle θ_S can be converted to the reflectance value of the corresponding surface element assuming that it has Lambertian characteristics and that it was irradiated by the sun at the standard zenith angle. $\theta_{SO} = 30°$ was assumed for this work. This radiometric calibration procedure is approximative only, as it does not account for direction-dependent influences of the atmosphere. As a consequence, the BRDF functions obtained in this analysis include the direction-dependent influence of the atmosphere.
For BRDF analysis, sample plots of different forest types and, for comparison, of other land cover types were selected and located on all images on which they are recorded. Details on a selection of sample plots are presented in chapter 4. The view azimuth angle and the view zenith angle were calculated for the centre of each plot considering the location of the plot in the image (distance to image centre in x and y), the principal distance of the camera and the orientation angles ω, φ, and κ (roll, pitch and yaw) of the image. Finally, the mean of the pixel values was extracted for each plot, for each image on which the plot appears, and for each spectral band. These pixel values were converted to spectral reflectance values as described above.
The models listed in chapter 3 were fitted to the data and compared with regard to their suitability to estimate the model parameters from the relatively small number of directional

reflectance data and to predict the directional reflectance characteristics of the plots. The Levenberg–Marquardt optimization algorithm was used for retrieving the model parameters. The MPFIT package by Markwardt (2009) programmed in IDL (RSI, 2004) was applied. It allows placing constraints on parameter values and fixing parameters. The model parameters were partly restricted to physically meaningful values.

The models were evaluated using the coefficient of determination (R^2) and the symmetric mean absolute percentage error (sMAPE) that is defined as

$$sMAPE = \frac{1}{n-p}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{(|y_i| - |\hat{y}_i|)/2} \cdot 100\% \quad , \quad (2)$$

where y_i is the measured reflectance, $\hat{y}_i$ is the modelled reflectance, n is the number of observations and p is the number of model parameters. The sMAPE was chosen instead of the RMSE, because it can be used to compare the error of sample plots that have different reflectance levels. The division by (n-p) instead of n incorporates model complexity in the error measure by penalizing a model for each model parameter.

3. BRDF MODELS

Over the past years, extensive efforts have been made to investigate the directional reflectance properties of surfaces. The result is a huge variety of models that predict the directional reflectance for all viewing geometries. The convention for denoting directions and angles is shown in Figure 2. The illumination and view directions are specified by 4 angles, the sun azimuth and zenith angle (ϕ_s and θ_s) as well as the view azimuth and zenith angle (ϕ_v and θ_v). The difference between the sun azimuth and the view azimuth angle is the relative azimuth ($\phi_r = \phi_s - \phi_v$).

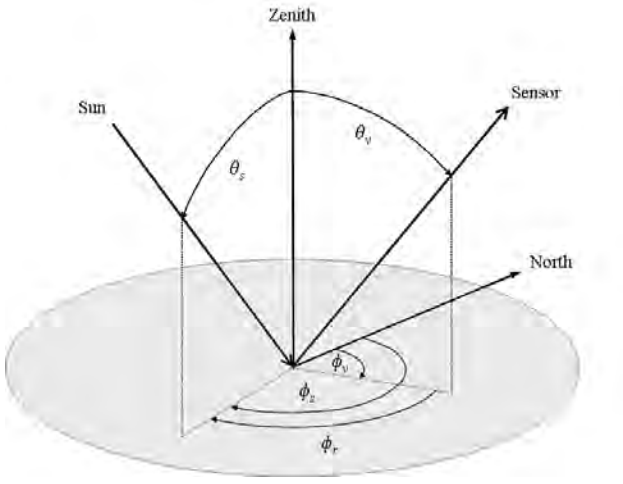


Figure 2. Convention for denoting directions and angles

The models employed in this study were selected according to the following criteria: physical, empirical and semi-empirical models should be included. The models should be suitable to describe the directional reflectance properties of vegetation in general and of forest canopies in particular and they should work for both visible and near-infrared light.

The following models were included in the comparison:

(1) Second-order polynomial (2DP): This model is a simple five-parameter second-order polynomial that is symmetric to the principal plane with two independent variables, the view zenith angle θ_v and the relative azimuth angle ϕ_r. The sun zenith angle θ_s is not considered.

(2) Walthall (WH): The Walthall model, derived by Walthall et al. (1985) and modified by Nilson and Kuusk (1989) to be reciprocal, is a purely empirical model for bare soil surfaces and vegetation canopy with four parameters. Liang and Strahler (1994) criticize that the model does not represent some important BRDF features such as the hot spot and amend the modified Walthall model by adding a hotspot component modelled by an exponential function with two additional parameters (WHL).

(3) Dymond (WAK): The model by Dymond et al. (2001) is a physical model that aims at reconstructing the bidirectional reflectance of homogeneous and closed vegetation canopies with randomly oriented leaves. Three parameters are used, one for the leaf reflectance and two to characterize the canopy phase function. The model is applicable to visible light and near-infrared light. It includes a term derived by Hapke (1981) to account for multiple scattering that is relevant in the near-infrared part of the spectrum. Two versions are proposed, WAK I and WAK II. The WAK II model rounds the hotspot as observed in the reflectance data of full size canopies, where there is both direct and diffuse illumination.

(4) Roujean (ROUJ): The model of Roujean et al. (1992) was the first kernel-driven model. It consists of a volume scattering kernel and a geometric-optical kernel. The volume scattering kernel is deduced from a formula by Ross (1981) and is suitable rather for optically thick than thin domains, i.e. for canopies with high values of leaf area index (LAI). Maignan et al. (2004) propose a modification of the volume scattering kernel to account for the hot spot more effectively (ROUJM). In contrast to the modification proposed by Chen and Cihlar (1997), no additional parameter is required.

(5) Ross-Thick/Li-Sparse (RTLS): This model combines the volume scattering kernel by Ross (Ross-Thick kernel) as described above with a geometric-optical kernel deduced from the model by Li and Strahler (1992), which considers the shadowing produced by randomly located trees with spheroid crowns on a Lambertian background. Wanner et al. (1995) deduced the Li-Sparse kernel, an approximation for sparse ensembles of such trees. Additionally to the three parameters usually used in kernel-driven models, there are two internal parameters describing the shape (oblate/round/prolate) and relative height of the tree crowns. The crown parameters in the best way describing the actual scene can be found by testing various sets of parameters. Both kernels, the Ross-Thick kernel and the Li-Sparse kernel, are implemented in AMBRALS, a modelling tool for the MODIS bidirectional reflectance and albedo products (Wanner et al., 1997).

(6) Rahman-Pinty-Verstraete (RPV): The RPV model (Rahman et al. 1993) is, in contrast to kernel-driven models, a multiplicative semi-empirical model. The base model uses three parameters (RPV3P). This model can be modified to a four-parameter model (RPV4P) that can improve modelled bidirectional reflectance values especially in the hotspot region when the hotspot effect is very pronounced. The RPV model is used for the processing of MISR surface products (surface BRF, albedo, LAI, FPAR) (Diner et al 2008).

4. DATA

The study is based on images taken with a Vexcel UltraCamD camera (Leberl et al., 2003). The multispectral bands without pansharpening were used. Parameters of the camera relevant for this study are listed in Table 3. The forward overlap was >80 % and the side overlap was 30 %.

Parameter	
FOV across track	37.5°
FOV along track	55°
Max. view zenith angle (diagonal)	31°
Pixel size (multispectral) at flying height of 3.900 m	75 cm
Radiometric resolution	12 bit

Table 3. Specifications of the UltraCamD digital aerial camera

The images were taken in the Rax-Schneeberg region in the south of Lower Austria in July 2006. For this study, six sample plots of different forest types and land cover types listed in Table 4 were selected. Criteria for the selection were moderate slope angles (<25 degree) and homogeneity in terms of tree species composition, development class and crown canopy closure. Furthermore, it was important that each plot is shown on images of two neighbouring flight lines to assure that the number of observations from different points of view is sufficient for the model fitting procedure. In Table 4, the area of each plot, the number of points of view and the range of the phase angle (i.e. the angle between the sun and view vector) is specified. The variation of the sun zenith angle is small for the entire image data set (27.1 to 30.7 degree), as the images were taken within a time interval of a few hours only.

SP		**Area [m²]**	**View directions**	**Phase angle**
1	Asphalt	130	16	8° - 48°
2	Grassland	160	13	16° - 57°
3	Young deciduous stand	350	12	5° - 38°
4	Old deciduous stand	700	11	5° - 50°
5	Young coniferous stand	650	10	6° - 52°
6	Old coniferous stand	2500	14	13° - 54°

Table 4. Sample plots

5. RESULTS

Due to limited space, only results for the near infrared band are given here.

As shown in Table 5, R^2 is quite high for all models and all sample plots. It lies between 0.852 (grassland, WAK I) and 0.999 (old coniferous stand, WHL). As R^2 is dependent on the variance of the observed values, here mainly caused by the direction-dependent reflectance, R^2 is generally lower for those sample plots that show smaller direction-dependent reflectance properties (asphalt, grassland). Thus, R^2 can only be compared for different models applied to the same sample plot. WHL achieves the highest R^2 for all sample plots but one (old deciduous stand), where WAK II outperforms the other models. However, it has to be considered that WHL is the model with the biggest number of model parameters (6). Obviously, the WHL model tends to overfitting due to the relatively small number of available observations (Table 4). Furthermore, the value of this purely empirical model is limited, because the model parameters cannot be applied to conditions (e.g. illumination geometry) that are outside the range covered by the observations, and because it does not permit the retrieval of biophysical parameters (e.g. LAI).

In contrast to R^2, the symmetric mean absolute percentage error (sMAPE; Equation 2) takes into account the model complexity, i.e. the number of parameters (Figure 8). However, it has to be noted that the different number of observations per sample plot still affects the comparison of different sample plots based on this measure.

The errors for the most successful models were between 3 and 6 %. The poorest results were achieved for SP5 (old coniferous stand).

For two sample plots (old coniferous stand, young coniferous stand) the difference in R^2 and sMAPE between ROUJ and ROUJM as well as between RPV3P and RPV4P is clearly visible. The hotspot effect is very pronounced for these test sites and it can be concluded that the hotspot-modification of these models is very effective. On the other hand, there is hardly any difference between RTLS and RTLSM, sMAPE is low for both models. The performance of ROUJ and RTLS, both kernel-driven models that apply the same volume scattering kernel, differs significantly. This indicates that the geometric-optical kernel of RTLS outperforms that of ROUJ, which is probably achieved by incorporating the two crown parameters. The simple second-order polynomial model performs surprisingly well except for the two sample plots with pronounced hotspot effect (young coniferous stand, old coniferous stand).

Figure 6 and 7 show examples for the BRDF (or, to be more specific, approximations to the BRDF, as noted above) at constant sun zenith angle for SP5 and SP6 modelled with the 3-parameter RPV model (RPV3P). The observations used to retrieve the model parameters are plotted (empty squares) and connected with the modelled values (filled squares) by a vertical line. The BRDF of the young coniferous stand is bowl-shaped with a slightly-developed hotspot effect, whereas the BRDF of the old coniferous stand is more bell-shaped and shows a pronounced increase of reflectance in the hotspot region. The reflectance modelled for the old coniferous stand is significantly lower than that of the young coniferous stand (apart from the hotspot region) which may be due to shadows.

The model parameters of the semi-empirical and physical models in most cases were plausible and touched the pre-set boundaries only in a few cases. Problems emerged if observations close to the hotspot were missing.

Model	**SP1** asphalt	**SP2** grass	**SP3** y dec	**SP4** old dec	**SP5** y con	**SP6** old con
2DP	0.914	0.856	0.938	0.990	0.958	0.975
WH	0.870	0.905	0.958	0.976	0.980	0.960
WHL	0.919	0.919	0.980	0.986	0.994	0.999
WAK I	0.903	0.852	0.965	0.989	0.975	0.995
WAK II	0.907	0.869	0.966	0.993	0.955	0.972
ROUJ	0.882	0.861	0.926	0.978	0.908	0.888
ROUJM	0.893	0.860	0.924	0.982	0.970	0.960
RTLS	0.897	0.870	0.927	0.985	0.989	0.997
RTLSM	0.893	0.867	0.927	0.983	0.989	0.998
RPV3P	0.909	0.857	0.921	0.981	0.944	0.968
RPV4P	0.909	0.858	0.927	0.985	0.980	0.998

Table 5. Coefficient of determination (R^2)

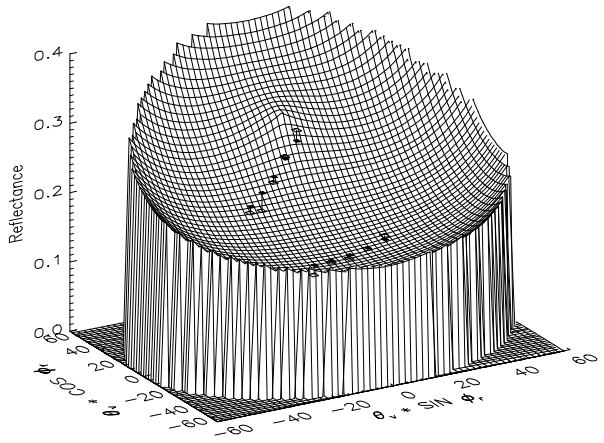


Figure 6. BRDF modelled with the 3-parameter RPV model (RPV3P) for a young coniferous stand (SP5), $\theta_s = 30^o$

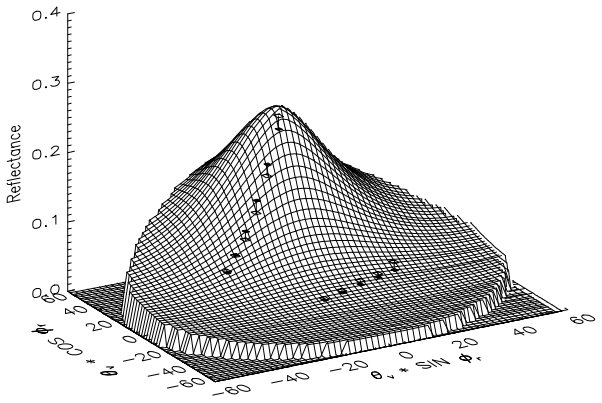


Figure 7. BRDF modelled with the 3-parameter RPV model (RPV3P) for an old coniferous stand (SP6), $\theta_s = 30^o$

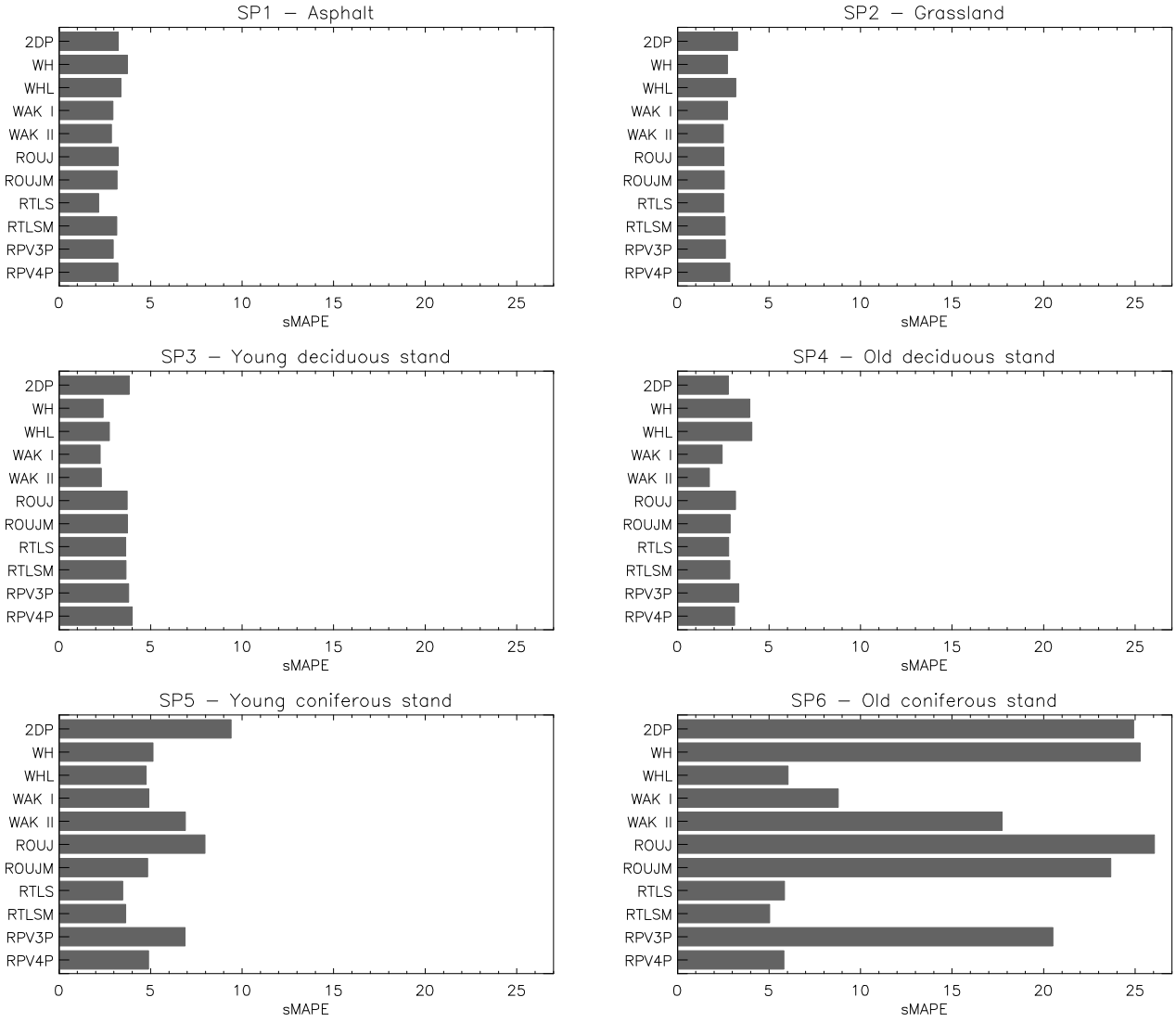


Figure 8. Symmetric mean absolute percentage error (sMAPE) for different BRDF-models and different land cover and forest types (SP1-SP6)

6. CONCLUSIONS

It has been shown that the parameters of BRDF models describing individual forest plots can be estimated from digital aerial photos taken with a frame camera with large forward and side overlap. The obtained results are plausible. Differences in the performance of the models for different forest plots can be explained taking into account the special assumptions on which the models are based and the special properties of the observed forest stands.

The following improvements and extensions are to be considered in future work:

(1) The number of test plots in the current study is small. More plots will be evaluated in order to deduce information on the variation of the derived model parameters for test plots of identical forest types, to assess the potential of BRDF characteristics obtained from aerial photos for discriminating forest types and determining forest parameters, and to verify this with independent test plots.

(2) The performance of the models in describing the dependence of the BRDF on the sun zenith angle will be studied using data sets with a higher variation of this parameter.

(3) The radiometric calibration will be improved by accounting for the non-Lambertian characteristics of the sample plots and the direction-dependent influence of the atmosphere.

(4) The analysis will be extended to pixel windows of different sizes down to full resolution provided by the camera (single pixels), in order to obtain more insight into the functioning and the performance of the models.

REFERENCES

Chen, J. M., Cihlar, J., 1997. A hotspot function in a simple bidirectional reflectance model for satellite applications. *Journal of Geophysical Research*, 102 (D22), pp. 25907-25913.

Diner, D. J., Martonchik, J. V., Borel, Ch., Gerstl, S. A. W., Gordon, H. R., Knyazikhin, Y., Myneni, R., Pinty, B. and Verstraete, M. M., 2008. Multi-angle Imaging Spectro-Radiometer - Level 2 Surface Retrieval Algorithm Theoretical Basis, Revision E. Jet Propulsion Laboratory, California Institute of Technology. http://www-misr.jpl.nasa.gov (accessed 25 Oct. 2009)

Dymond, J. R., Shepherd, J. D., Qi, J., 2001. A simple physical model of vegetation reflectance for standardising optical satellite imagery. *Remote Sensing of Environment*, 77 (2), pp. 230-239.

Dymond, J. R., Trotter, C. M., 1997. Directional reflectance of vegetation measured by a calibrated digital camera. *Applied Optics*. 36 (18), pp. 4314-4319.

Hapke, B., 1981. Bidirectional reflectance spectroscopy, 1. Theory. *Journal of Geophysical Research*, 86, pp. 3039-3054.

Leberl, F., Perko, R., Gruber, M., Ponticelli, M., 2003. The UltraCam large format aerial digital camera system. *Proceedings of the American Society for Photogrammetry and Remote Sensing Annual Convention*, Anchorage, Alaska, 5-9 May 2003, published on CD, no page numbering.

Li, X., Strahler, A. H., 1992. Geometric-Optical Bidirectional Reflectance Modeling of the discrete crown vegetation canopy: effect of crown shape and mutual shadowing. *IEEE Transactions on Geoscience and Remote Sensing*, 30 (2), pp. 276-292.

Liang, S., Strahler, A. H., 1994. Retrieval of surface BRDF from multiangle remotely sensed data. *Remote Sensing of Environment*, 50, pp. 18-30.

Maignan, F., Bréon, F. M., Lacaze, R., 2004. Bidirectional reflectance of Earth targets: evaluation of analytical models using a large set of spaceborne measurements with emphasis on the hot spot. *Remote Sensing of Environment*, 90 (2), pp. 210-220.

Markwardt, C. B., 2009. Non-linear least squares fitting in IDL with MPFIT. In: Bohlender, D. A., Durand, D., Dowler, P. (eds.), *Proceedings of Astronomical Data Analysis Software and Systems XVIII*, Quebec, Canada, ASP Conference Series, Vol. 411, Astronomical Society of the Pacific: San Francisco, pp. 251-254.

Nilson, T., Kuusk, A., 1989. A reflectance model for the homogeneous plant canopy and its inversion. *Remote Sensing of Environment*, 27, pp 157-167.

Rahman, H., Pinty, B., Verstraete, M. M., 1993. Coupled surface-atmosphere reflectance (CSAR) model - 2: Semiempirical surface model usable with NOAA advanced very high resolution radiometer data," *Journal of Geophysical Research*, 98, pp. 20791–20801.

Ross, J., 1981. *The radiation regime and architecture of plant stands*. Dr. W. Junk Publishers. The Hague.

Roujean, J.-L., Leroy, M., Deschamps, P.-Y., 1992. A bidirectional reflectance model of the earth's surface for the correction of remote sensing data. *Journal of Geophysical Research*, 97 (D18), pp. 20455-20468.

RSI, 2004. IDL – Interactive Data Language. Boulder, CO

Schaepman-Strub, G., Schaepman, M. E, Painter, T. H., Dangel, S., Martonchik, J. V., 2006. Reflectance quantities in optical remote sensing – definitions and case studies. *Remote Sensing of Environment*, 103, pp. 27-42.

Smith, G. M., Milton, E. J., 1999. The use of the empirical line method to calibrate remotely sensed data to reflectance. *Int. Journal of Remote Sensing*, 20 (13), pp. 2653-2662.

Walthall, C. L., Norman, J. M., Welles, J. M., Campbell, G., Blad, B. L. 1985. Simple equation to approximate the bi-directional reflectance from vegetation canopies and bare soil surfaces. *Applied Optics*, 24, pp. 383-387.

Wanner, W., Li, X., Strahler, A. H. 1995. On the derivation of kernels for kernel-driven models of bidirectional reflectance. *Journal of Geophysical Research*, 100 (D10), pp. 21077-21089.

Wanner, W., Strahler, A. H., Hu, B., Lewis, P., Muller, J.-P., Li, X., Schaaf, C. L. B., Barnsley, M. J. (1997). Gobal retrieval of bidirectional reflectance and albedo over land from EOS MODIS and MISR data: Theory and algorithm. *Journal of Geophysical Research*, 102 (17), pp. 17143-17161.

LAND COVER IDENTIFICATION USING POLARIMETRIC SAR IMAGES

A. Kourgli [a, *], M. Ouarzeddine [a], Y. Oukil [b], A. Belhadj-Aissa [a]

[a] Image Processing laboratory, FEI, U.S.T.H.B., 16111 Bab-ezzouar, Algiers, Algeria
[b] Geography Department, ENS, Bouzareah, Algiers, Algeria
a_kourgli@lycos.com

KEY WORDS: SAR, Classification, Texture, Modelling, distributed

ABSTRACT:

Synthetic Aperture Radar (SAR) has been proven to be a powerful earth observation tool. Due to its sensitivity to vegetation, its orientations and various land-covers, SAR polarimetry has the potential to become a principle mean for crop and land-cover classification. A variety of polarimetric classification algorithms have been proposed in the literature for segmentation and/or classification of polarimetric SAR images into classes reflecting canonical scattering processes and/or some statistical properties. However, classification based on polarimetric data alone does not provide sufficient sensitivity for the separation of some classes such as forests. The use of other kinds of characteristics like texture provides better sensitivity for class separation. In this paper, we wish to address this issue, testing and comparing some polarimetric SAR classification algorithms using texture. Such an analysis will allow us to evaluate the importance of texture considering and to prove if the chosen texture model parameters describe, also, physical properties of the targets. Thus, the proposed approach is compared with the Wishart classifier showing interesting results. The test area used is the Oberpfaffenhofen in Munich and the SAR images are acquired in the P band.

1. INTRODUCTION

Synthetic Aperture Radar (SAR) has been proven to be a powerful earth observation tool. In many remote sensing applications important additional information can be derived from multipolarised imagery. The number of studies and applications involving polarimetric SAR data is increasing steadily. Radar polarimetry has for long been known as a powerful method for soil moisture and surface roughness identification, sea ice detection and delineation of vegetation and land cover (Ferro-Famil, 2003; Macri, 2003; Hoekman, 2003; Lee, 2004; Martini, 2004; Wakabayashi, 2004; Alberga, 2007; Liang, 2008). Many features such as intensities, coherency matrix, correlation and phase differences have been used in various classification experiments.

The first algorithms developed for classification of polarimetric SAR images have ignored the spatial information (texture) and used the Whishart distribution as the basis of the classification scheme. But, th last decade, several research papers revealed the contribution of texture in polarimetric classification improvement. Thus, some authors used texture for classification without a decomposition method. Yu and Acton (2000) presented a partitioning scheme using an initial texture segmentation based on watershed algorithm. Ersahin *et al.* (2004) have proposed a neural network unsupervised classification scheme using covariance matrix parameters and texture features derived from gray level co-occurrence matrices. Recently, some authors proposed to employ texture features calculated from polarimetric data after decomposition. So, Beaulieu and Touzi (Beaulieu, 2004) introduced a segmentation algorithm that takes into account texture information where the K-Wishart distribution is used to model textured areas. Rodionova (2007) demonstrated that textural features defined in every scattering categories of Freeman and Durden decomposition make better object discrimination of SAR polarimetric images. In (Khan, 2007), good classification results have been achieved using neural network with a feature set including undecimated wavelet, transform-based features and texture features along with nonlinear features and a partial set from the elements of the coherence matrix. Liang (2008) also investigated the performance of different texture features using neural network classifier. Bombruno (2008) demonstrated that the use of an appropriate texture distribution is useful to segment textured PolSAR images. Dabboor *et al.* (2008), also, combined the textural features in each scattering category obtained from the Freeman-Durden decomposition with the number of the scattering mechanisms from the entropy calculated from the Cloude-Pottier decomposition. in order to perform the segmentation process. Zhang *et al.* (Zhang, 2009) combined the scattering powers of MCSM (Multiple-Component Scattering Model) and selected texture features from Gray-level co-occurrence matrices using SVM (Support Vector Machine) classifier and neural network (Zhang, 2009).

The objective of this paper is to evaluate the performance of texture modelling of polarimetric SAR images in land-cover classification by two steps scheme: the first step is Cloude and Pottier decomposition and the second one is markovian textural classification applied on decomposed images. For evaluation purpose, the result is compared to a classification obtained using Wishart classifier.

2. POLARIMETRIC SAR CLASSIFICATION

Many supervised and unsupervised classification methods have been proposed, such as methods based on the maximum likelihood (ML), artificial neural networks (NN), support vector machines (SVM), fuzzy methods, etc. An usual approach is to classify polarimetric SAR images based on the inherent characteristics of physical scattering mechanisms using decomposition theorems (Cloude, 1996; Touzi, 2004). Several decomposition techniques were proposed. These techniques are

* Corresponding author.

based on three principal approaches known as coherent methods, Huynen decomposition (Huynen, 1970) and non coherent methods. These methods split the scattering matrix into the sum of elementary scattering matrices, each one defining a deterministic scattering mechanism (Touzi, 2007). As we are dealing with texture which is a neighbourhood property, we choose to compare the textural classifier to a classifier based on incoherent decomposition. Indeed, non coherent decompositions permit to take into account the context, thus neighbourhood. For this purpose, we used whishart classifier. In fact, the characteristic decomposition of the Hermitian target coherency matrix allowed Cloude and Pottier to derive key parameters, such as the scattering type α and the entropy H, which have become standard tools for target scattering classification and for physical parameter extraction from polarimetric SAR data.

2.1 Textural Classifier

Image texture, defined as a function of the spatial variation in pixel intensities, is useful in a variety of applications and has been a subject of intensive study by many researchers. The texture parameter is extremely important for radar imagery interpretation, especially in terrain mapping. In fact, Radar image depends heavily on the scattering of ground objects and its textures/structures strongly vary with different objects. Many models have been employed in texture analysis including autoregressive model, Markov random fields (MRF), Gaussian random fields, Gibbs random fields, World model, wavelet model, multichannel Gabor model, fractal model, etc (Chellapa, 1993; Bader, 1995; Arivazaghan, 2003). In this study, we use a non parametric Markovian model that has been successfully applied to SAR image classification (Kourgli, 2009) and adapt its formulation to polarimetric images.

2.1.1 Texture model: Markov theory states that each pixel in an image has an independent local spatial property characterized by its surrounding neighbours. A discrete Markov field {X} is defined on a 2-D lattice S with a neighbourhood structure N_S. Its global properties (Π) are controlled by means of local properties which are defined by local conditional probabilities (Derin, 1987; Li, 2000) The Markov property describes local conditional dependence of pixels image as:

$$\begin{aligned} &\forall\, x \in \Omega, s \in S, \\ &\Pi_s\left(X_S = x_S \,/\, X_r = x_r, r \in S, r \neq s\right) \\ &= P\left(X_S = x_S \,/\, X_r = x_r, r \in N_S\right) \end{aligned} \tag{1}$$

Another attractive property of an MRF is that, by the Hammersley-Clifford theorem (Hammersley, 1971) an MRF can be characterized by a global Gibbs distribution that is usually defined with respect to cliques. A clique C is a particular spatial configuration of pixels, in which all its members are statistically dependent of each other:

$$P(x_s \,/\, x_r, r \in N_s) = \frac{\exp(-U(x))}{Z_s} \tag{2}$$

Where $U(x) = -\sum_{c \in C_S} V_c(x)$ is called energy and is obtained by summing potential functions $V_C(x)$ while Z_S is a normalizing constant. A probability model is usually specified by a parametric probability distribution. The model is to be identified, in order to find best values for unknown parameters of the model for a given training texture. Due to usually complex mathematical form of texture distribution and because almost natural textures are quasi-stationary, parametric models fail to model them. Hence, we adopted an energy formulation defined by a Neighbourhood Likeness Measure (NLM) which is estimated between the neighbourhood N_S and all the neighbourhoods N_Y contained in the texture sample Y.

This measure is given by:

$$\begin{aligned} U(x) &= NLM\,(N_S, N_{y \in Y}) \\ U(x) &= \frac{1}{card\,(N_y)} \sum_{y_r \in N_y, x_r \in N_x} (y_r - x_r)^2 \end{aligned} \tag{3}$$

Texture modelling can be performed in order to reproduce natural textures, but it can also be used as a tool for a classification or for a segmentation purpose.

2.1.2 Classification: The texture segmentation problem is the labelling of pixels in a lattice to one of texture classes, based on a texture model and the observed intensity field. Bayesian approaches, where maximum *a posteriori* (MAP) estimation is usually involved to image segmentation, have been proven efficient. In addition, we adopted the assumption presented in (Bouman, 1994) which states that the joint probability of a pixel in a window F_S can be approximated by the product of the neighbourhood probabilities over this window:

$$\Pi\ (x_s,\ s \in F_s) \approx \prod_{s \in F_s} P_s\,(x_s \,/\, x_r, r \in N_s) \tag{4}$$

As a combined use of physical scattering characteristics and statistical properties for terrain classification is desirable, we apply textural classification on Pauli decomposed vector.
The Pauli vector, for a full polarimetric data, is given by:

$$\vec{K}_P = \frac{1}{\sqrt{2}} \begin{bmatrix} S_{HH} + S_{VV} \\ S_{HH} - S_{VV} \\ 2S_{HV} \end{bmatrix} \tag{5}$$

The first element of the vector expresses odd bounce scatterer type such as the sphere, the plane surface or reflectors of trihedral. The second one is related to a dihedral scatterers or double isotropic bounce and the third element is related to horizontal and a cross polarising associated to the diffuse scattering or volume scattering.

Using the probability defined above and Pauli decomposition, we performed a classification scheme which is described as follows:

- Select from each texture image a sample.
- Scan the mosaic texture using a window sized F_S with a step of one pixel in the row and column directions, and calculate the joint probabilities defined by equation (4) for each sample in the three decomposed images
- The central pixel of the window considered will be assigned to the class maximizing the joint probabilities calculated.

2.2 Wishart classifier

In this work we are using the non coherent decomposition proposed by Cloude and Pottier. It is based on the coherency matrix calculated using the Pauli basis and given by (Cloude and Pottier, 1997):

$$[T] = \left\langle \vec{k}_p \cdot \vec{k}_P^{\,T} \right\rangle \tag{6}$$

This is a multilook 3x3 positive semi-definite hermitian coherency matrix where the superscript T denotes the matrix transpose, and < > indicates multilook averaging. The $\sqrt{2}$ on the term is to ensure consistency in the span (total power) computation. The eigenvectors and eigenvalues of the coherency matrix [T] can be calculated to generate a diagonal form of the coherency matrix which can be physically interpreted as statistical independence between a set of target vectors.

2.2.1 Entropy, alpha and anisotropy: The eigenvalues of [T] have direct physical significance in terms of the components of scattered power into a set of orthogonal unitary scattering mechanisms given by the eigenvectors of [T], which for radar backscatter form the columns of a 3 x 3 unitary matrix. Hence, we can write an arbitrary coherency matrix in the form (Papathanassiou, 1999):

$$\langle [T_3] \rangle = [U_3][\Sigma][U_3]^{-1} = \sum_{i=1}^{3} \lambda_i \vec{u}_i \vec{u}_i^{\,*T} \tag{7}$$

where $[\sum]$ is a 3x3 diagonal matrix with nonnegative real elements:

$$[\Sigma] = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \tag{8}$$

$[U_3]$= [u1 u2 u3] is a unitary matrix, where u1, u2 and u3 are the three unit orthogonal eigenvectors

After eigen vector decomposition of the coherency matrix the entropy *H*, which is a measure of the randomness of the scattering process, is deduced from the eigen values as:

$$H = -\sum_{i=1}^{3} P_i \log_3 (P_i) \tag{9}$$

where P_i are the probabilities obtained from the eigenvalues λ_i :

$$P_i = \frac{\lambda_i}{\sum_{i=1}^{3} \lambda_i} \tag{10}$$

The entropy *H* represents the randomness of the scattering. *H* = 0 indicates a single scattering mechanism (isotropic scattering) while *H* = 1 indicates a random mixture of scattering mechanisms with equal probability and hence a depolarising target.
The parameter α is indicative of the average or dominant scattering mechanism. It describes the dominance of the scattering mechanism in terms of volume, double bounce or surface scattering types. The α angle is obtained from the α_i angles of each of the eigen vectors as follow:

$$\overline{\alpha} = \sum_{i=1}^{3} P_i \alpha_i \tag{11}$$

The anisotropy *A* is a parameter complementary to the entropy. It measures the relative scattering of the second and the third eigenvalues of the eigen-decomposition. It is given by:

$$A = \frac{\lambda_2 - \lambda_3}{\lambda_2 + \lambda_3} \tag{12}$$

If the pair H −α plotted on a plane then they are confined to a finite zone. This plane is subdivided into eight zones characterizing different classes corresponding to different scattering mechanisms.

2.2.2 Classification: The basic scattering mechanism of each pixel of a polarimetric SAR image can be identified by comparing its entropy and parameters to fixed thresholds. The different class boundaries, in the H-alpha plane, have been determined so as to discriminate surface reflection (SR), volume diffusion (VD) and double bounce reflection (DB) along the α axis and low, medium and high degree of randomness along the entropy axis. When the anisotropy parameter is introduced, it allows the possibility to distinguish different clusters where the centers belong to the same H −α partition (Ouarzeddine, 2007).

The eight classes resulted from the H−α decompositions are used as training sets for the initialization of the unsupervised Wishart classifier. For a coherency matrix <Ti> of a pixel i of a multilook image (L-looks) knowing the class ωi, the Wishart complex distribution is given by:

$$p\left(\langle [T_i] \rangle | \omega_m\right) = \frac{L^{Lq} \left| \langle [T_{3i}] \rangle \right|^{L-q}}{K(L,q) \left| V_m \right|^{L}} \exp\left(-Ltr\left(V_m^{-1} \langle [T_{3i}] \rangle\right)\right) \tag{13}$$

where $\Sigma_m = E\left(\langle[T_i]\rangle \middle| \langle[T_i]\rangle \in \omega_m\right) = \frac{1}{N_m}\sum_{i=1}^{N_m}\langle[T_i]\rangle$

N_m is the pixel number of ω_m
$K(L,q)\,|$ is the factor of standardization given by :

$$K(L,q) = \pi^{q(q-1)/2}\prod_{i=1}^{q}\Gamma(L-i+1) \quad (14)$$

Where q=3 in the case of reciprocity (i.e., Shv=Svh), | | and tr (.) indicate the determinant and the trace of the matrix respectively, and Γ(.) is the gamma function.
A probabilistic measurement of the distance between the a coherence matrix of an unspecified pixel <Ti>, and the average coherence matrix Σ_m of the class candidate ω_m , is obtained using:

$$d\left(\langle[T_i]\rangle, \Sigma_m\right) = \ln\left(|\Sigma_m|\right) + tr\left(\Sigma_m^{-1}\langle[T_i]\rangle\right) \quad (15)$$

Mathematically, each coherency matrix of an individual pixel is assigned with the most likely class ω_m with the minimal distance, if and only if :

$$d\left(\langle[T_i]\rangle, \Sigma_m\right) \le d\left(\langle[T_i]\rangle, \Sigma_j\right) \quad (16)$$

for all $\omega_m=\omega_n$.

The clusters of the first itération are used us a training set for the second itération until a consequent result is obtained.

3. DATA USED

The data is related to the site of Obepfaffenhofen and is captured on May 2000. It is covering the southern part of Munich, Germany and an area that embraces about 1 Km^2 was chosen in this study. The radar used is the Aeos1 of the Ex private Aerosensing GmBH company. An airborne radar with the P band (72 cm). The SAR images are acquired in fully polarimetric mode and are of important value to supply information on the terrain type. An area of 600×600 pixels has been extracted for this study (Figure 1).

Figure 1: A colour composite of the test site using the Pauli basis.

4. CLASSIFICATION RESULTS

From Wishat classification result, we selected six samples sized 12×12 pixels, three forest types (magenta, yellow and cyan), two for base area (blue and green), and the last represents built up area (red). First, we considered each image from Pauli vector decomposition apart and performed textural classification in each scattering category. The result obtained for each decomposed image is shown in Figure 2.a, 2.b and 2.c, while Figure 2.d is a colour composite of the three probability images before labelling.

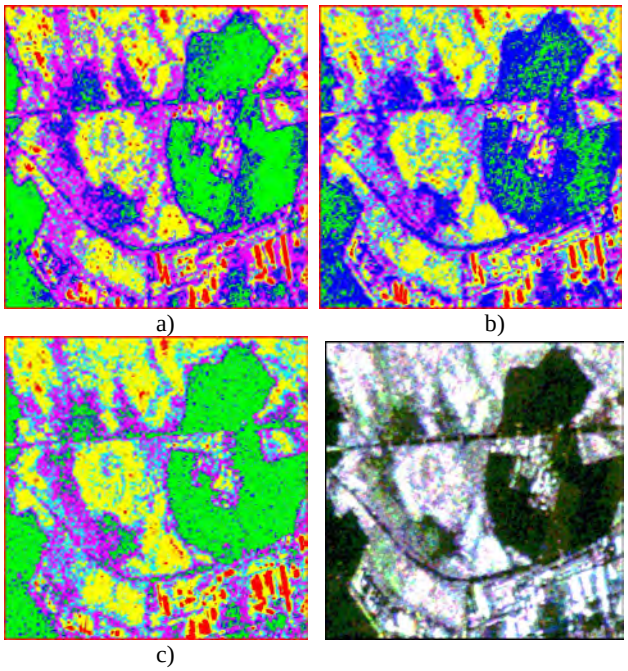


Figure 2: The textural classification using Pauli images (a,b and c) and their colour composite (d)

The different classes are well segmented in all the three scattering channels. Indeed, the result obtained is interesting since all the classes are almost correctly recognized, however, some confusion occurs at the boundaries of different classes. These preliminary results show that the model texture is sensitive to the polarization (the classification results are different for each scattering type) and that the different scattering mechanisms have also been discriminated. The composite colour (Figure 2.d) of the different segmented decompositions provides a first classification that is interesting.
In a second step, we classified the image on the basis of the three Pauli images simultaneously maximizing probability over the three images. The results are illustrated in Figure 3. We have reduced the classification result of Wishart classifier to the same number of classes. The polarimetric textural classification (Figure 3.b) appears consistent with the result obtained from Wishat classifier (Figure 3.c). We notice a large similarity for base area (in blue and in green) and built-up area (in red).

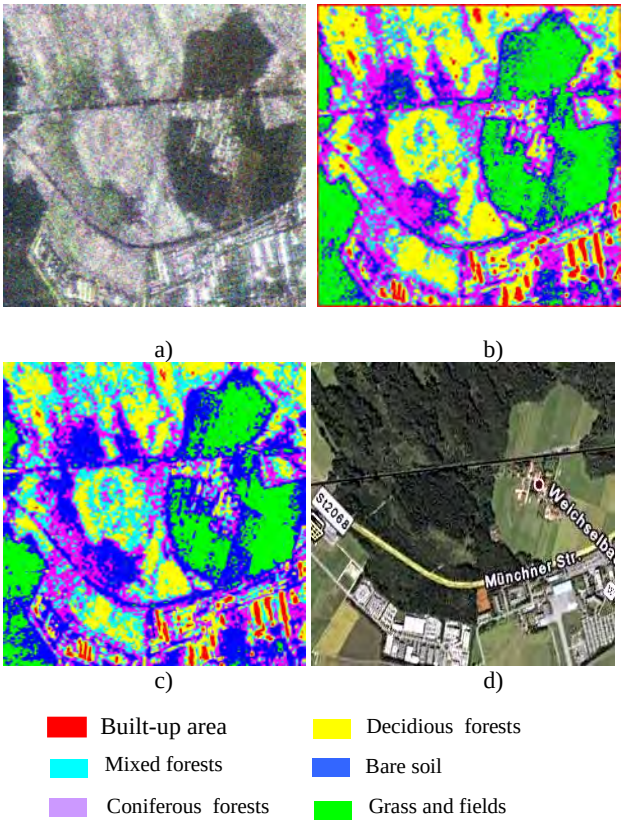


Figure 3: A colour composite of the test site (a) using the Pauli basis decomposition, polarimetric classification (b : textural classifcation, c : Wishart-A-α classification) and an image extracted from google earth software of the test site

For a comparison purpose and to check the efficiency of our textural classification, we computed the classification percentage for each class (table 4) taking the Wishart classification result as a reference.

Classes	Color+rate	
Built-up	Red	89.53
Mixed forests	Cyan	44.45
Deciduous forests	Yellow	86.86
Bare soil	Blue	47.73
Grass and fields	Green	91.56
Coniferous forests	magenta	68.28

Table 4: Classification percentage

While showing good rates for three classes (>86%), textural classification performs less better for the other samples. The weak rates obtained for bare soil (in blue) and mixed forests (in cyan) could be justified by a bad localization of the corresponding samples used in textural classification process. Furthermore, we give at Figure 3.d an optical image (Google Earth) of the same area. We can see that buit-up area (in red) has well identified by the textural classification, but some confusion occurs between decidious forests (in yellow) and built-up area. This gives a clear indication of how the classification algorithm has performed. Thus, textural modelling performed in every scattering type of Pauli decomposition makes good object discrimination of SAR polarimetric images if samples are correctly chosen. This can be done, for example, by using H −α partition.

5. CONCLUSION

The different polarisations respond in different ways to the orientation and shape of the objects from which scattering takes place. It is obviously important to have tools that make full use of this information. The main purpose of this paper was to test object separability performance by using texture modelling in each scattering type. As the information in the fully polarimetric data can not be completely represented by one single feature, the combination of different polarimetric features incorporating spatial information seems to give interesting results. Indeed, textural modelling of polarimetric images provides a possibility to separate different classes and proves the texture features polarization dependence. It was shown that texture significantly contributed to land cover discrimination where backscattering coefficient could fail. Thus, our results confirm previous findings that texture incorporating can improve polarimetric images interpretation and help in land cover identification.

REFERENCES

Alberga, V., 2007. A study of land cover classification using polarimetric SAR parameters. *International Journal of Remote Sensing*, 28(17), pp. 3851-3870.

Bader, D. A., Jaja, J., Chellapa, R., 1995. Scalable data parallel algorithms for texture synthesis using gibbs random fields. *IEEE Transactions on Image Processing*, 4 (10), pp. 1456-1460.

Beaulieu, J-M., Touzi, R., 2004. Segmentation of textured polarimetric SAR scenes by likelihood approximation. *IEEE Transactions on Geoscience and Remote Sensing*, 42(10), pp. 2063-2072.

Bombrun, L. and Beaulieu, J. M., 2008. Fisher distribution for texture modeling of polarimetric SAR data. *IEEE Geoscience And Remote Sensing Letters*, 5(3), pp. 512-516.

Bouman, C. A., Shapiro, M., 1994. A multiscale random field model for bayesian image segmentation. *IEEE Transactions on Image processing*, 3(2), pp.162-177.

Cloude S.R. and Pottier E., 1996. A review of target decomposition theorems in radar polarimetry. *IEEE Transactions on Geoscience and Remote Sensing*, 34(2), pp.498-518.

Dabboor, M., Karathanassi, V., Braun, A., 2008. Land cover classification of ALOS polarimetric SAR data. IGARSS08, July 6-11, Boston, USA.

Derin, H., Elliott, H., 1987. Modelling And Segmentation Of Noisy And Textures Images Using Gibbs Random Fields. *IEEE Transactions of Pattern Analysis And Machine Intelligence*, Pami-9(1), pp. 39-55.

Ersahin K., Scheuchl B. and Cumming I., 2004. Incorporating texture information into polarimetric radar classification using neural networks. IGARSS‘ 2004 *Proceedings IEEE International Geoscience and Remote Sensing Symposium, 2004*, 1.

Ferro-Famil, L. Pottier, E. Lee, J. S.., 2003. Unsupervised classification of natural scenes from polarimetric interferometric

SAR data, in *Frontiers of Remote Sensing Information Processing*, Edtior: C.H. Chen, World Scientific99.

Hammersley, J. H., Clifford, P., 1971. Markov field on finite graphs and attices, Technical Report, Unpublished.

Huynen, J. R., 1970. Phenomenological theory of radar targets. Technical Report, University of Technology, Delft, The Netherlands.

Hoekman D. H., Martin A. M., 2003. A new polarimetric classification approach evaluated for agricultural crops. *IEEE Transactions on Geoscience and Remote Sensing*, 41(12), pp.2881-2889.

Khan, K. U., Yang, J., 2007. Polarimetric synthetic aperture Radar image classification by a hybrid method, *Tsinghua Science and Technology*,12(1), pp. 97-104.

Kourgli, A., Belhadj-Aissa, A., Oukil, A., 2009. SAR image classification using textural modelling. Radar 2009, 12-16 October, Bordeaux, France.

Lee, J.-S., Grunes, M. R., Pottier, E. and Ferro-Famil, L.,2004 Unsupervised terrain classification preserving polarimetric scattering characteristics. *IEEE Transactions on Geoscience and Remote Sensing*, 42, pp. 722-731.

Li, S. Z., 2000. Modeling image analysis problems using random markov field", in *Handbook of Statistics*, 20. Elsevier Science, pp. 1-43.

Liang, G., Yifang, B., 2008. Investigating the performance of sar polarimetric features in land-cover classification. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.* 36(B6b), pp. 317-321.

Macri Pellizzeri, T., 2003. Classification of polarimetric SAR images of suburban areas using joint annealed segmentation and H/A/α polarimetric decomposition. *ISPRS Journal of photogrammetry and remote sensing*, 58 (1-2), pp. 55-70.

Martini.A., Ferro-Famil.L., Pottier.E., 2004. Multi-frequency polarimetric snow discrimination in Alpine areas, *IGARSS'04 Proceedings*, 6, pp. 3684-3687.

Ouarzeddine, M., Souissi, B., Belhadj-Aissa, A., 2007. Unsupervised classification using wishart classifier. Workshop of POLinSAR 2007, 22 - 26 January 2007,ESA-ESRIN, Frascati, Italy.

Papathanassiou, K.P., 1999. Polarimetric SAR interferometry, PhD_Thesis the faculty of natural sciences, department of physics, technical university Graz, Austria.

Rodionova, N. V., 2007. A combined use of decomposition and texture for terrain classification of fully polarimetric sar images. . Workshop of POLinSAR 2007, 22 - 26 January 2007,ESA-ESRIN, Frascati, Italy.

Touzi, R., Boerner, W. M., Lee, J. S. and Lueneburg, E., 2004. A review of polarimetry in the context of synthetic aperture radar: concepts and information extraction. *Canadian Journal of Remote Sensing*, 30 (3), pp. 380-407.

Touzi, R., 2007. Target Scattering Decomposition in Terms of Roll-Invariant Target Parameters. *IEEE transaction on Geoscience and. Remote Sensing*, 45 (1), pp. 73-84.

Wakabayashi.H., Matsuoka.T., Nakamura.K., Nishio.F., 2004. Polarimetric Characteristics of sea ice in the sea of Okhotsk observed by airborne L-band SAR. *IEEE Transactions on Geoscience and Remote Sensing*, 42 (11), pp. 2412-2425.

Yu, Y., Acton, S. T., 2000. Polarimetric SAR image segmentation using texture partioning and statistical analysis, Proceedings of *IEEE International Conference on Image Processing*, Vancouver, Canada, Sept. 10-13, 2000., pp. 677-680.

Zhang, L., Zou, B., Zhang, J., Zhang, Y., 2009. Classification of polarimetric SAR image based on support vector machine using multiple-component scattering model and texture features. *Eurasip Journal on Advaces in Signal processing*. Hindawi.

Zhang, Y-D., Wu, L-N., Wei, G., 2009. A new classifier for polarimetric SAR images. *Progress In Electromagnetics Research, PIER* 94, pp. 83-104.

ANALYSIS OF SPATIOTEMPORAL DIFFERENCE OF NDVI IN AN ARID COAL MINING REGION USING REMOTE SENSING

S. Lei [a, *], Z. Bian [a]

[a] Jiangsu Key Laboratory of Resources and Environmental Information Engineering, China University of Mining and Technology, Xuzhou,221116, Jiangsu Province, China
lsgang@126.com

Key words: Spatial, Statistics, Vegetation, Landsat, Analysis

Abstract:

Different resolution of remote sensing images will give rise to different perspectives and spatial characteristics. The objective of this study is to compare the spatiotemporal difference of the vegetation index extracted from TM and MODIS images by time series analysis and spatial statistics, and find the relationship among between the vegetation, climate factors, coal mining etc..The study area is located at an arid mine area, where the mining activities and ecology reconstruction is ongoing. It is found that the MODIS-NDVI (monthly or 16 days) products can provide results close to the NDVI derived from atmosphere corrected TM images. Time series analysis found that the monthly NDVI, rainfall and temperature are consistently subject to annual periodical rhythm under the impacts of coal mining. And there is a significant correlation between NDVI and rainfall & temperature in the arid mine area. However, MODIS-NDVI (1 km) is not suitable for spatial statistics for the study area of 3200 km^2, because of the coarse spatial resolution. NDVI-TM (30 m) or NDVI-MODIS (250 m) are feasible for spatial statistics at this study area. Higher value of NDVI is accompanied by higher spatial variation of NDVI with a squared correlation coefficient (R^2 =0.6983). It is probably because the natural arid landform was damaged by human activities, e.g. vegetation construction and industry.

1. INTRODUCTION

Vegetation changes play an important role in the environmental processes, and also is a sensitive indicator for environmental and global changes (Van Wijngaarden, 1991). The vegetation monitoring can provide useful clues concerning our changing environment and help natural resource management. Traditional method to monitor the vegetation is by field investigation. It is low efficiency and high labor demanding, especially for large scale area, and impossible to conduct continuously investigation. Nowadays, advanced Remote Sensing (RS) is a powerful monitoring tool for its convenience and high efficiency. Thereby, it has been widely employed to monitor the vegetation changes (e.g. Justice & Hiernaux, 1986; Townshend & Justice, 1986; Hobbs, 1995; Al-Bakri & Taylor, 2003).

Vegetation coverage, leaf area index and vegetation index are the main indices of vegetation information. However vegetation coverage and leaf area index are often obtained based on vegetation index. Vegetation index is a simple numerical indicator, which can be derived directly from RS image. For example, there are ARVI (atmospherically resistant vegetation index, Kaufman & Tanre, 1992), SAVI (soil adjusted vegetation index, Huete, 1988), NDVI (Normalized Difference Vegetation Index, Rouse et al 1974), EVI (Enhanced Vegetation Index, Liu & Huete 1995), etc. NDVI is one of the most important and commonly used vegetation indexes, defined as equation (1).

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)} \quad (1)$$

where RED is the reflectance in the red channel and NIR is the reflectance in the near-infrared channel. The RED and NIR band contain more than 90% of vegetation information. (Baret et al, 1989)

The NDVI can be derived from many different kinds of RS images, e.g. Landsat, Spot, MODIS, NOAA/AVHRR, etc. with different spatial and temporal resolution. For example, the NDVI derived from TM image is 30 m spatial resolution and 16 days temporal resolution. The MODIS (Moderate Resolution Imaging Spectroradiometer) is at lower spatial resolution, but with a daily temporal resolution. Previous researches showed that multi-temporal NDVI images are useful for analyzing spatial vegetation pattern and for assessing vegetation dynamics (e.g. Justice & Hiernaux, 1986; Townshend & Justice, 1986). Time-series analyses of satellite data enable the observation of seasonal and annual trends of vegetation cover (Vicente et al, 2004). Su et al. (2001) pointed that the results or conclusions for the same area may change with different spatial resolution of research scale. And it is important to understand the effect of different resolution images on the analysis of spatial variation. Suitable resolution of RS image is needed for spatial analysis at a given scale.

Therefore, the objective of this study is to find the spatial and temporal variation of vegetation under the influence of local arid climate, vegetation reconstruction and mining activities, at a mine area with arid and semi-arid climate and located in north of China.

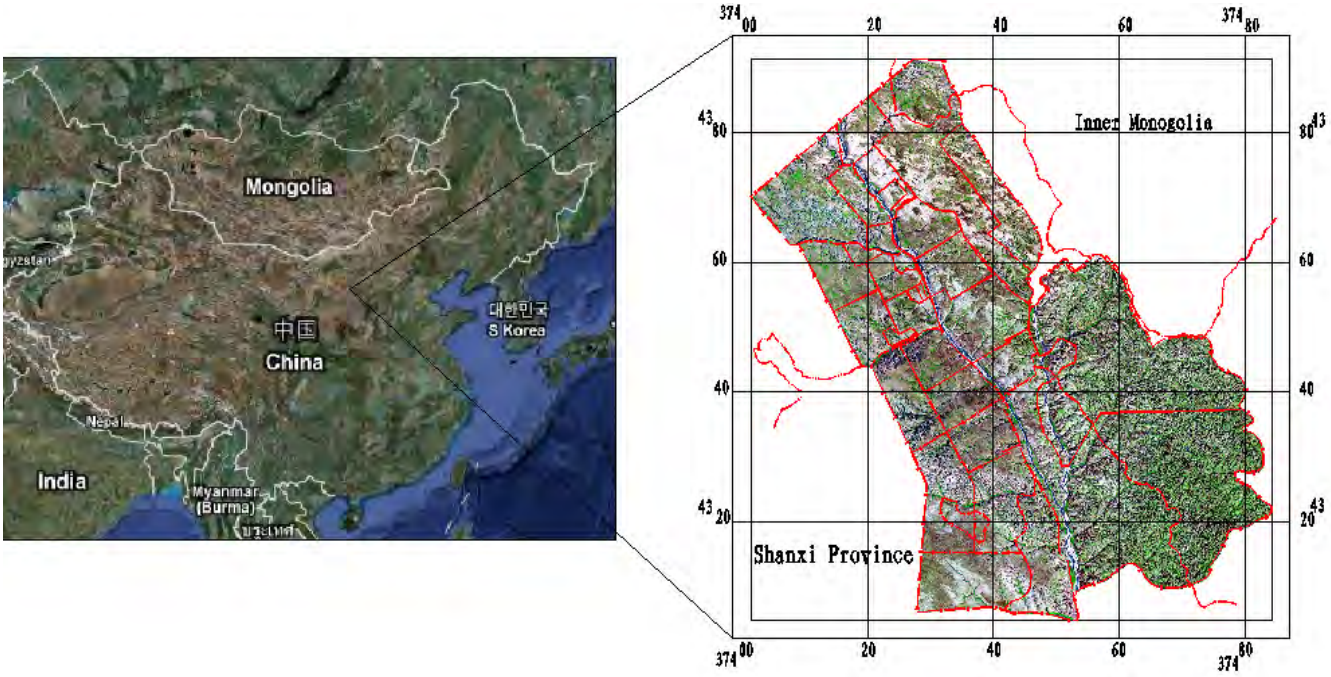
2. METHODOLOGY

2.1 Case study area

The experimental area is Shendong coal mine area located at the border of Shanxi province and Inner Monogolia (figure 1). The total area size is about 3200 km^2; and the elevation is between 1000 m to 1500 m. The average annual rainfall is about 436.7 mm, about 70% of which happens in July, August and September (Cui et al., 2001). And the groundwater is deep below the surface and intermittent. The surface is aeolian landform with sparse vegetation which is typical sandy land vegetation. There is also some grasses and planted vegetation, such as poplar and Salix psammopbylla. In general, this area

* Corresponding author

has a fragile and unstable ecological environment and low-level



social development, characterized by arid and semi-arid climate.

Figure 1. The location of Shendong coal mine area

2.2 Dataset

(1) Landsat TM/ETM+ satellite data

Three Landsat TM images acquired on August 24th 2000, August 6th 2002 and July 5th 2005 were processed. The atmosphere effect was corrected by using FLAASH model (Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes) (Berk *et al*., 1998) in ENVI 4.3. The NDVI can be calculated based on equation (1), then the mean NDVI of the study area can be obtained by statistics.

(2) MODIS-NDVI products

MODIS launched in December 1999 has high temporal resolution (at least twice per day,) and can be free downloaded (Lu and Zhao, 2005). There are several types of MODIS-NDVI products with different spatial and temporal resolution produced by Earth Observation System (EOS). The spatial resolution of MODIS-NDVI-16days-250m is 250 m. The NDVI images were converted into the projection of WGS84, UTM from Sinusoidal projection, and tailored by the boundary of study area with ENVI4.3. Further study found that the MODIS-NDVI-monthly-1km imagery has the similar average NDVI values as the MODIS-NDVI-16days-250m imagery. In this study, 67 MODIS-NDVI-monthly-1km and several MODIS-NDVI-16days-250m images from January 2000 to August 2005 were processed.

2.3 Spatiotemporal statistics

Temporal changes of mean NDVI were derived from monthly MODIS vegetation products to reflect the vegetation change varying with time. Moreover, the correlations between the NDVI and the climates factors, e.g. monthly rainfall and air temperature were analyzed to determine the influences of climate factors on the vegetation in this arid mining region.

A semivariogram analysis was used to describe the spatial variance and spatial structure of the NDVI. The semivariance statistic was calculated as one half of the average squared difference between data values at pairs of points a given distance apart (Journel and Huijbregts, 1978). It is calculated as:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \qquad (2)$$

where γ(h) is the empirical semivariance for the distance h, N(h) the number of points separated by the distance h, and $Z(X_i)$ is the NDVI at location X_i. In this case, the X_i is the pixel. The γ(h) was then plotted against the lag distance h, yielding the empirical semivariogram, which characterizes the spatial variability of NDVI as a function of distance. For predictions, the empirical semivariogram is converted to a theoretic one by fitting a statistical model. For example, equation (3) is the exponential model.

$$\gamma(h) = C_0 + C\left[1 - \exp\left(-\frac{|h|}{a}\right)\right] = C_0 + C[\exp\ (|h|)] \qquad (3)$$

where C_0 is the nugget, the semivariance at a distance 0. a is the distance at which the semivariogram levels off. C_0+ C = Sill .Sill is the value at which the semivariogram levels off. Higher sill indicates higher spatial variation of NDVI. The semivariance can be calculated in ENVI 4.3, and the empirical semivariogram was regressed in SPSS (Statistical Package for the Social Science).

3. RESULTS AND DISCUSSION

3.1 Temporal change of NDVI

Table 2 shows the NDVI derived from Landsat TM and MODIS images at the same period. It is noticed that the NDVI with FLAASH correction provides a much closer and more reasonable fit with the MODIS-NDVI than the uncorrected NDVI. Therefore, it is necessary to remove the atmosphere effect before calculation of NDVI. Moreover, the difference

between the corrected NDVI and MODIS-NDVI is due to the different methods of NDVI acquisition. Specifically, the NDVI of Landsat is for one day; but MODIS-NDVI is the maximum NDVI over a period, e.g. 16 days or one month.

Time	uncorrected	FLAASH corrected	MODIS-NDVI Monthly	MODIS-NDVI 16days
2000-8-24	0.0549	0.3480	0.2542	0.2926
2002-8-6	0.0534	0.3251	0.3346	0.3288
2005-7-5	0.1098	0.2432	0.2894	0.2823

Table 2. Comparison of different NDVI at the same period

Time series of NDVI-MODIS of the study area from January 2000 to August 2005 are presented in figure 3. It is found that the vegetation of the whole Shendong mine area has been improved in recent years, under the influence of climate, mining and ecological restoration. The improvement of vegetation was also found by Hu & Chen (2008). Figure 3 also shows that the NDVI, rainfall and air temperature are well subject to annual periodical rhythm. And there are significant correlations between monthly NDVI and two climate factors. The correlation coefficient of NDVI with rainfall and temperature is 0.760, and 0.84, respectively under the impact of coal mining.

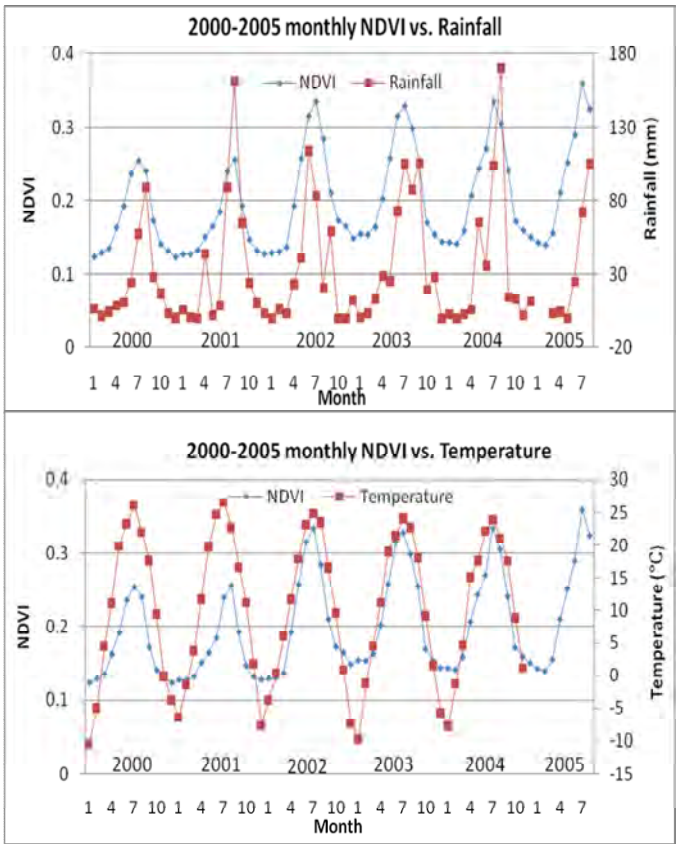


Figure 3. Comparison between monthly NDVI-MODIS and temperature & rainfall

3.2 Spatial characteristics

Different spatial resolution of images will result in different spatial semivariance of NDVI. So it is important to choose the right resolution for spatial characteristics analysis. Figure 4 presents the difference of the semivariance of NDVI-MODIS (250 m and 1 km). For the 250 m NDVI, it is revealed that when the distance is at 1750 m (250 m*7 pixels) the semivariogram levels off. However, for the 1 km NDVI, the semivariogram does not level off, even the distance is up to 50 000 m (1 km*50 pixels). It indicates that NDVI-MODIS with 250 m resolution is feasible for spatial variance analysis in the study area. But the NDVI with 1 km resolution is not suitable for spatial analysis of the study area.

Moreover, the vegetation variance will be reflected by the semivariance. Figure 4 also shows that the semivariance of NDVI in the December (0.1113) is lower than that in summer (0.1932). That is because the study area is in the arid area with sparse vegetation. In winter, most of the vegetation are wilted or died, and the ground is almost bare. Therefore, the homogeneity of ground surface decreases the spatial variation.

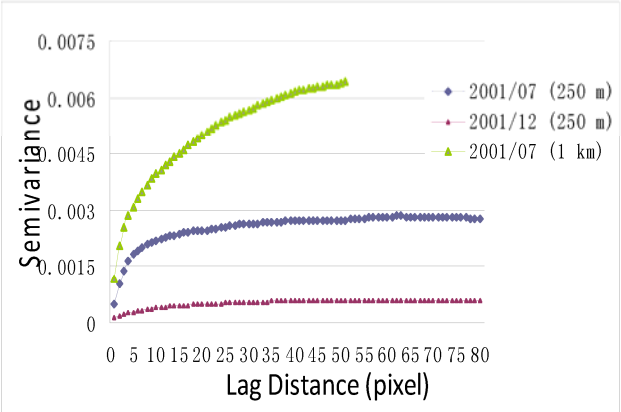


Figure 4. Semivariogram comparison of NDVI-250 m and 1 km. X axis is the lag distance of pixel, the unit of NDVI-1km is 1 km, NDVI-250m is 250 m)

The spatial characteristics obtained from different spatial resolution (30 m, 250 m, 1 km) of NDVI were analyzed and compared by semivariance analysis. The semivariance of NDVI was presented in figure 4 and figure 5. It was found that the NDVI of 2002 and 2003 in July are of higher semivariance, because of the higher value of NDVI. Detailed analysis can be made by using a empirical semivariance model. The exponential model was determined as optimal empirical semivariance model, the key parameters of which were provided in table 6.

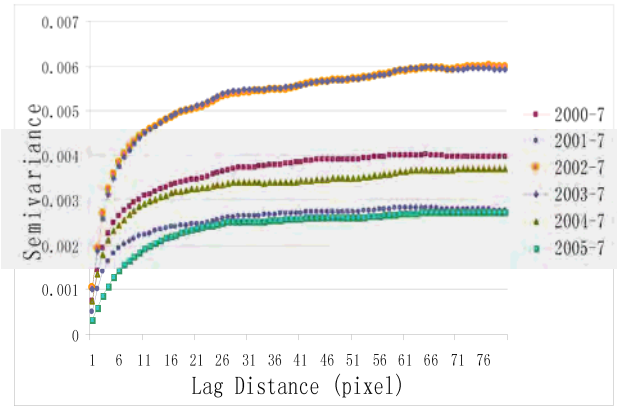


Figure 5. Semivariogram of NDVI-MODIS (250 m), in July, from 2000 to 2005; X axis is the lag distance, 250 m per pixel

TM	2000-8-24	2005-7-5	MODIS	2000/7	2001/7
C_o	0.00208	0.00149	C_o	0.00101	0.00063
C	0.00629	0.00407	C	0.00290	0.00212
a	11.94064	8.84545	a	9.29724	7.97316
$C+C_o$	0.00836	0.00556	$C+C_o$	0.00391	0.00275
NDVI	0.3480	0.2432	NDVI	0.2423	0.1932
MODIS	2002/7	2003/7	2004/7	2005/7	2001/12
C_o	0.00151	0.00130	0.00080	0.00020	0.00011
C	0.00429	0.00449	0.00274	0.00249	0.00048
a	10.14961	9.51552	7.71630	9.55438	12.70900
$C+C_o$	0.00580	0.00579	0.00354	0.00269	0.00059
NDVI	0.3288	0.3403	0.2766	0.2823	0.1113

Table 6. Key parameters of exponential semivariance model for NDVI-TM (30 m) and NDVI-MODIS (250 m), a is pixel distance. For the exponential semivariance model, when the distance is 3*a, the semivariogram levels off.

A high correlation (R^2=0.6983) between NDVI and Sill ($C+C_o$) was found, shown in figure 7. It indicates that the higher value of NDVI is accompanied by higher spatial variation of NDVI. For example, the mean NDVI in July 2003 is 0.3403, and the $C+C_o$ is 0.00579. But the $C+C_o$ is only 0.00275 for July 2001, when the NDVI is 0.1932. It is probably because the sparse vegetation coverage of the study area is lower than 50%. And the wide natural bare landform is damaged by different kinds of human activities, which will increase the spatial variation, e.g. vegetation construction or mining industry, etc.

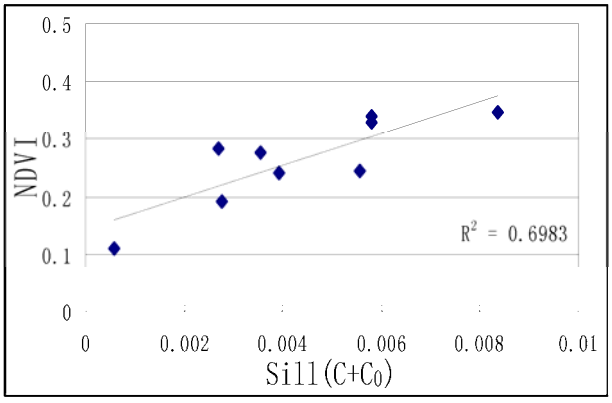


Figure 7. Correlation between NDVI and Sill ($C+C_o$)

4. CONCLUSIONS

MODIS-NDVI and Landsat-TM images were processed to study the spatiotemporal variance of vegetation in an arid mining area. It is found that MODIS-NDVI (monthly or 16 days) products can provide results close to the NDVI derived from atmosphere corrected TM images. Time series analysis found that the monthly NDVI, rainfall and temperature are consistently subject to annual periodical rhythm. The arid climate variables are still the dominating factors, instead of underground coal mining. Furthermore, there is a high correlation between NDVI and rainfall & temperature in arid mine area, where the mining activities and ecology reconstruction is ongoing. Therefore, the NDVI-MODIS (1 km) is good for time series analysis at large scale, e.g. region scale or global scale.

Different resolution of images, e.g. TM (30 m), MODIS (250 m or 1 km) will give rise to different spatial characteristics. However, 1 km MODIS-NDVI is not suitable for spatial statistics for the study area, which is about 3200 km^2, because of the coarse spatial resolution. NDVI-TM (30 m) or NDVI-MODIS (250 m) are feasible for spatial statistics at local scale. Higher NDVI is accompanied by higher spatial variation. A high positive correlation (R^2 =0.6983) between the value of NDVI and the spatial variation of vegetation (sill) was found. It is probably because the homogeneous natural arid landform was damaged by human activities, e.g. vegetation construction and industry.

REFERENCES

Al-Bakri, J.T. and Taylor, J.C. 2003. Application of NOAA AVHRR for monitoring vegetation conditions and biomass in Jordan. Journal of Arid Environments, 54, pp.579-593.

Baret, F., Guyot, G. and Major, D.J., 1989. TSAVI: A vegetation index which minimizes soil brightness effects on LAI and APAR estimation. Proceedings of the 12th Canadian Symposium on Remote Sensing. Vancouver ,Canada,1989,pp. 1355-1358.

Berk, A., L.S. Bernstein, G.P. Anderson, P.K. Acharya, D.C. Robertson, J.H. Chetwynd, and S.M. Adler-Golden., 1998. MODTRAN Cloud and Multiple Scattering Upgrades with Application to AVIRIS, Remote Sensing of Environment, 65, pp.367-375.

Cui, Y.L, Shao J.L., & Han, S.P., 2001.Ecological Environment Adjustment by Groundwater in Northwest China. Earth Science Frontiers, 8, pp.191-196.

Hu, Z.Q. and Chen, T., 2008. EARDAS aided extraction of vegetation fraction from remote sensing information in coalmine area based-a case study of Shenfu coalmine. Journal of Northwest Forestry University, (232), pp.164-167.

Huete, A.R., 1988. A soil adjusted vegetation index (SAVI). Remote Sensing of Environment, (25), pp.295 - 309.

Journel, A.G., and Ch.J. Huijbregts. 1978. Mining geostatistics. Academic Press, New York.

Justice, C.O. & Hiernaux, P.H.Y., 1986. Monitoring the grasslands of the Sahal using NOAA AVHRR data: Niger 1983. International Journal of Remote Sensing, 7, pp. 1475–1497.

Kaufman, Y.J. and Tanre, D. 1992. Atmospherically resistant vegetation index (ARVI) for EOS – MODIS. IEEE Trans Geosci Remote Sensing, (30), pp.261 - 270.

Liu, H.Q., and Huete, 1995. A Feedback Based Modification of the NDVI to Minimize Canopy Background and Atmospheric Noise . IEEE Trans Geosci Remote Sensing, (33), pp.457 - 465.

Lu, L.P., Zhao, C.Y. 2005. Study on the spatial change of the different desert vegetation Index based on MODIS Data--A case study of Sangong River Basin. ARID LAND GEOGRAPHY, 28(3), pp.381-384.

Rouse, J.W., 1974. Monitoring the vernal advancement of retrogradation of natural vegetation. NASA/GSFC, Type III, Final Report, Greenbelt, MD, pp. 371.

Su, L.H., Li X.W. and Huang Y.X., 2001. AN REVIEW ON SCALE IN REMOTE SENSING. Advance In Earth Sciences, 4, pp.544-548.

Townshend, J.R.G. and Justice, C.O., 1986. Analysis of dynamics of African vegetation using the NDVI. International Journal of Remote Sensing, 7, pp. 1224–1242.

Van Wijngaarden, W. 1991. The green cover of the Earth: a dynamic resource in a changing environment. ITC Journal, 3, pp.113–121.

Vicente-Serrano, S. M., Lasanta, T. , Romo, A., 2004. Analysis of Spatial and Temporal Evolution of Vegetation Cover in the Spanish Central Pyrenees: Role of Human Management Environmental Management ,34, pp. 802-818.

ACKNOWLEDGEMENTS

The Project Sponsored by the Scientific Research Foundation of Jiangsu Key Laboratory of Resources and Environmental Information Engineering (Grant No. JS200904), the 111 project with approved No. B07028, and the project for Ph.D. program with approved No. 20060290511.

A FRAMEWORK OF POLARIMETRIC SAR FILTER BASED ON INDEPENDENCY OF INTENSITY AND POLARIMETRIC INFORMATION

P.Li [a], S. Deng [a, *], J. Zhang [b], G. Huang [b], and Z. Zhao [b]

[a] State Key Lab. for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, Hubei Province, China – pxli@lmars.whu.edu.cn, dsp410327@yahoo.com.cn
[b] Chinese Academy of Surveying and Mapping, No. 28th, Lianhuachi West Road, Haidian District 100830, Beijing, China - zhangjx@casm.ac.cn, (huang.guoman, zhengzhaochina)@163.com,

KEY WORDS: Radar, SAR, Polarization, Imagery, Land cover, filter, speckle

ABSTRACT:

The usual polarimetric speckle filters optimally combine the polarization channels into a single-channle image or only restore the radiometric information. The phase differences and the polarization channel correlation coefficients are lost. Though simple boxcar filter and extended Lee filter for covariance matrix can be used to preserve polarimetric information without changing the data form, the results are not satisfying enough. In this study, we proposed a new framework for polarimetric SAR filter in which phase and intensity information are filtered independently based on the independence of intensity and polarimetric information. Firstly, polarimetric covariance matrix is divided into the product of a scalar z and a matrix C_0. The sum of diagonal elements in C_0 is 1 and z denotes the sum of diagonal elements in covariance matrix which is also called total power or SPAN. SPAN z and C_0 include intensity or texture information and polarimetric information respectively which both are relatively independent. Secondly, SPAN image is filtered as singly polarized imagery, for example using Lee filter. To preserve the polarimetric information, the boxcar filter is applied to C_0, which can be considered as averaging pixels in a moving window using equal weights for each pixel. Also, the filtered C_0 must be normalized by dividing each element by the sum of its diagonal elements. Finally, the filtered result is obtained by multiplying the filtered SPAN and filtered C_0. Experiment results show that the proposed filter is valid, and polarimetric information is well preserved. And compared with extended Lee filter, the performance of the proposed method is better.

1. INTRODUCTION

Synthetic aperture radar (SAR) is an active radar imaging system working on microwave band and can be operated under all weather condition without daylight. Thus it is now widely used as a tool of earth observation. However, for its coherent imaging mechanism, SAR images are inevitably corrupted by speckle noises (Lee, 2009). Speckle noise makes it more difficult to recognize terrain target and classify land cover than other remote sensing images. It is necessary and useful to suppress speckle noise before application. In recent years, many methods have been developed to reduce speckles in SAR images. For example, Lee, Frost, Kuan filters are effective in speckle suppression (Lee, 1980; Frost, 1982; Kuan, 1987). In order to preserve edges, lines or point targets, many new adaptive methods were proposed, and worked well. However, most of them are only fit for singly polarized real images.

On the other hand with the development of airborne and spaceborne radar, users can obtain fully polarized SAR (PolSAR) images more conveniently. The same as non-fully polarized images, the application of PolSAR data is also restricted for its speckle. There are several speckle filters which are often applied to PolSAR images. Polarimetric whitening filter proposed by Novak is often used to detect targets (Novak, 1990). And Multi-texture polarimetric whitening filter was developed by G.Q. Liu (Liu, 1995). The disadvantage of these methods is that the data form is changed to single- or multi-real channels. So the phase difference disappears, and polarimetric information is not well preserved. Another filter frequently used is optimal weighting filter developed by J.S. Lee. It requires that the input data is real by which it is impossible to preserve phase difference. In order to keep the data form, assembling averaging and many improved assembling averaging method such as sigma filter are applied. Lee filter is also extended to polarimetric images. Other preferable filter is intensity driven adaptive neighbourhood filter (Vasile, 2006). Using these methods, speckle is well suppressed when data form is kept. Nevertheless the result is unstable in heterogeneous areas; especially there are some point targets in the neighbourhood. The filtered covariance matrix of centre pixel is often predominated by these point targets. So is the polarimetric information. That is the problem which exists in most filters developed before.

In this study, we assume that all information in PolSAR images is made of intensity and polarimetric information or phase difference expressed by a covariance matrix of which the sum of diagonal elements is 1. Both are relatively independent. The filtering is based on this assumption model. The intensity image and polarimetric information are filtered respectively. Finally, combining the filtered two parts, we obtain the speckle suppressed image with the same data form as the input image.

2. FRAMEWORK OF POLARIMETRIC FILTER

2.1 Model of PolSAR data

A fully polarized synthetic aperture radar measures both amplitude and phase of the four polarized backscattered signal

* Corresponding author.

HH, HV, VH VV, producing a 2×2 complex matrix or Sinclair matrix:

$$S = \begin{bmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{bmatrix} \quad (1)$$

where $S_{HH}, S_{HV}, S_{VH}, S_{VV}$ = complex backscattering coefficients of HH, HV, VH, VV polarization channel respectively.

In the reciprocal backscattering case, $S_{HV} = S_{VH}$. The Sinclair matrix can be transformed to a three dimensional vector:

$$y = \begin{bmatrix} S_{HH} & S_{HV} & S_{VV} \end{bmatrix}^T \quad (2)$$

where T = the transpose of a vector.

In order to describe the polarimetric information more explicitly, the polarimetric data is usually converted to a 3×3 covariance matrix from polarimetric vector (Lee, 2009):

$$C = yy^H \quad (3)$$

where C = covariance matrix
Y = backscattering vector
H = the conjugate transpose of a complex vector.

The PolSAR data expressed by (1) contains two kinds of information. One is the information of the three intensities which is the diagonal elements of covariance matrix in (3)(3), and the other one is the information of the three phase differences between the three polarized channels which are included in the non-diagonal elements of the covariance matrix. Actually there are two independent differences because:

$$\arg(C_{12}) + \arg(C_{23}) = \arg(C_{13}) \quad (4)$$

where arg(·) = argument of a complex number
C_{12} = complex covariance of S_{HH} and S_{HV}
C_{23} = complex covariance of S_{HV} and S_{VV}
C_{13} = complex covariance of S_{HH} and S_{VV}.

We consider that the PolSAR imagery in the form of covariance matrix is made of two parts:

$$C = zC_0 \quad (5)$$

where z = total power of backscattering coefficients (SPAN)
C_0 = normalized covariance matrix (or polarimetric information part, or phase differences part)
C = observation of covariance matrix.

The first part z is a scalar, denoting the intensity or the total power and is the sum of the diagonal elements of C as $z = C_{11} + C_{22} + C_{33}$.

The total power z also called SPAN includes the information of texture. The other part C_0 is a matrix denoting the phase differences between the three polarimetric channels, including polarimetric information. The sum of diagonal elements in C_0 is 1. When each element of covariance matrix is divided by z, C_0 is obtained.

As depicted above, backscattering from a target is made of two parts, the intensity and polarimetric information. And because different targets may have different intensity, and the polarimetric information may be different, that is to say, the two aspects have no definite correlations, and both can be considered independent from each other. The filter proposed in this paper is base on this product model. The filtered covariance matrix is also made up by two part, SPAN and normalized covariance matrix.

2.2 Filter Framework Based on Independency of Intensity and Polarimetric Information

In unfiltered PolSAR images, not only intensity is degraded by speckle, but also the polarimetric information is degraded by speckle. In section 2.1, we know that intensity and polarimetric information or phase differences in PolSAR data are independent. Therefore, the two parts could be filtered separately.

2.2.1 Filter for intensity image. For intensity image, the distribution was discussed in many articles in detail, and many filters are developed based on the probability density function of clutters. Lee, Kuan, gamma and sigma filter are very useful. In order to preserve details, for example lines, edges and point targets, refined filters were developed, and the performance was improved. In a word, all filters which could be applied to singly polarized SAR intensity images could also be used for the filtering of the intensity part.

In this study, J.S. Lee filter was selected for the filtering of intensity part for its good adaptability. Lee filter is based on Statistic in a moving window:

$$\hat{x} = (1-b)\bar{Z} + bZ \quad (6)$$

Where $\hat{x}$ = the estimation of true value for backscattering intensity
Z = the observation in window centre
$\bar{Z}$ = the mean of the intensity in the moving window.

Coefficient b in (6) can be calculated (Lee, 1980):

$$b = \frac{1 - \sigma_v^2 \bar{Z}^2 / \mathrm{var}(Z)}{1 + \sigma_v^2} \quad (7)$$

where var(Z) = the variation of Z in the moving window
σ_v^2 = the variation of ideal speckle based on multiplicative model. For single look complex

(SLC) SAR images, $\sigma_v^2 = 1$.

2.2.2 Filter for polarimetric information. Polarimetric information is mainly included by the covariance matrix. There are two primary factors. One is phase difference between different polarized channels. The other one is the modulus of complex covariance of different polarized channels. Actually, both are also represented in the normalized covariance matrix C_0 of which the SPAN is considered as 1.

To make the description above more understandable, SPAN could be considered as a weight. In C_0, the weight is set to 1 while in original covariance matrix the weight is the total power, SPAN. Apparently, It is more reasonable to assuming that the weight of each pixel for polarimetric information are equal to 1 than equal to SPAN if polarimetric information is primarily concerned. Therefore, Normalized covariance matrix is preferable comparing with original covariance matrix when polarimetric information preservation is considered as one of the most indexes of performance for PolSAR filter.

In homogeneous areas, each matrix for a resolution unit represents polarimetric information with equal weight. So the boxcar filter has top-priority. However, actually for SAR images, pixels in the neighbourhood are not always homogeneous. If we use origin covariance directly again as the polarimetric boxcar filter and extended J.S. Lee filter, the intensity plays a real role as weight actually, and the mechanism of strong targets such as points is magnified. Therefore polarimetric information is confused, and errors may be brought to the results. That is why the improvement was limited although many adaptive techniques were introduced.

But when the representation of polarimetric information, covariance matrix is normalized and averaged with equal weight, the influence of pixels with mechanisms different from centre pixel is restricted. In other words, polarimetric mechanisms for different pixels are assigned the same weight in the boxcar filter of normalized polarimetric covariance matrix.

Thus, in homogeneous or heterogeneous areas, the normalized boxcar filter for polarimetric information in the data form of covariance is better. In this step, we only need to apply boxcar filter to the normalized covariance matrix. In addition, after boxcar filtering, the filtered covariance matrix $\hat{C}_0$ should also be normalized divided by the sum of its diagonal elements:

$$\hat{C}_0 = \bar{C}_0 / (\bar{C}_{011} + \bar{C}_{022} + \bar{C}_{033}) \tag{8}$$

where $\bar{C}_0$ = the mean of C_0 in the moving window
$\hat{C}_0$ = estimation of normalized covariance matrix
$\bar{C}_{0ii}$ = the diagonal elements of C_0 at Colum i and Row i.

So that, each pixel of filtered normalized covariance matrix has the same intensity, or the same weight as the input normalized covariance matrix.

2.2.3 Combination for the filtered two parts. According to the model represented in section 2.1, multiplying the filtered intensity $\hat{x}$ by filtered normalized covariance matrix $\hat{C}_0$, we get the final filtering result for PolSAR images:

$$\hat{C} = \hat{x}\hat{C}_0 \tag{9}$$

where $\hat{C}$ = estimation of polarimetric covariance matrix
$\hat{x}$ = estimation of SPAN (or intensity)
$\hat{C}_0$ = estimation of normalized covariance matrix (or polarimetric information).

2.3 General Steps for Filter based on Intensity and Polarimetric Information

Generally speaking, there are three primary steps in the framework for filter based on the independence of the intensity and polarimetric information which are described in Figure 1. Now we discuss the sequences of the general three steps as follows:

1) Calculate intensity and normalized covariance matrix from Sinclair matrix according to (5), as the top of Figure 1.
2) Apply filters to intensity image and normalized covariance matrix respectively according to (6) and (8). For Intensity image, filters such as Lee, Kuan, sigma and gamma are compatible. For Normalized covariance matrix, in order to preserver polarimetric information, the mean filter with equal weight is the best. This step is described in the middle of Figure 1.
3) Combine the two filtered results obtained in step 2, and then get the final filtered PolSAR images as (9) and the bottom of Figure 1 which also contains the texture and polarimetric information with the same data form as the input data.

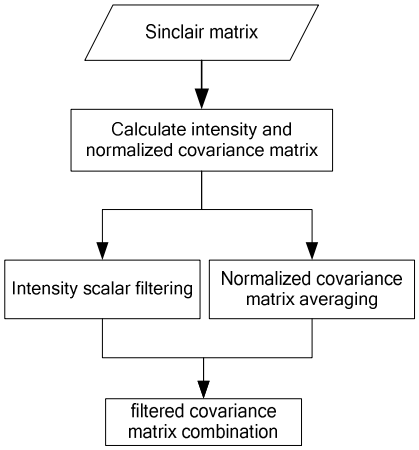


Figure 1. The follow chart of filtering based on intensity and phase difference model

2.4 Some Improvement for the Filter in the Framework

Although there is some disadvantages for polarimetric information preservation in the filter proposed above. Some improvements are possible and necessary.

In order to preserve textures or other details, techniques which are used in singly polarized intensity images to preserve textures or edges can also be extended to this filter framework. One method is changing the pixel set in neighbourhood or the moving window shape adaptively in which statistics are calculated. Taking refined Lee filter for example, firstly we need to determine which one of the eight non-square sub-windows the centre pixel belongs to, and then we calculate the statistics in the non-square sub-windows, finally we obtain the filtered covariance matrix as (9).

In addition, point target preservation is another index for the performance of a filter. A const false alarming rate (CFAR) method can be applied to the SPAN image to detect point targets firstly. Then in the process of filtering, the covariance matrix of point target is remained the same as that of the input. At the same time, point target is excluded from the neighbourhood where the local mean and variance is calculated in both intensity and normalized covariance matrix.

SPAN is usually assumed to have a gamma distribution, the CFAR used in this paper is based on this distribution. The threshold T is obtained by solving the equation (Gigly, 2002):

$$p_{fa} = 1 - \int_0^T \frac{1}{\Gamma(n)} \left(\frac{n}{\sigma}\right)^n t^{n-1} \exp(-\frac{nt}{\sigma}) dt \qquad (10)$$

where p_{fa} = the false alarming rate,
n = the number of looks, for SLC data n=1
σ = normalized standard derivation in a moving window, $\sigma = \sqrt{\mathrm{var}(Z)/E(Z)}$.

For a const false alarming rate, for example $p_{fa} = 0.005$, the centre pixel of which the intensity is larger than T is determined as point targets.

3. EXPERIMENTAL RESULTS

3.1 Data Description

The single look C band Radarsat 2 PolSAR image used in this study was acquired in 2007 over some suburb. The image is in the form of Sinclair matrix with the size of 512 by 512 pixels with a spatial resolution of about 10 meters.

For PolSAR imagery, Pauli decomposition is usually used to produce a colour image because this decomposition can be used to visualize PolSAR data based on the three primary scattering mechanisms (Lee, 2009). Pauli decomposition can be described as:

$$k = \left[S_{HH} + S_{VV} \quad S_{HH} - S_{VV} \quad 2S_{HV} \right] \qquad (11)$$

where k = coherent polarimetric vector
$S_{HH}, S_{HV}, S_{VH}, S_{VV}$ = complex backscattering coefficients of the four polarized channel respectively.

Using the three elements of the polarimetric vector as red, green and blue channel, for example, Red = $|S_{HH}+S_{VV}|$, Green = $|S_{HH}-S_{VV}|$, Blue = $|S_{HV}|$ where $|\cdot|$ denotes calculating the modulus of a complex, we can obtain a colour image based on Pauli decomposition.

Figure 2 shows the original PolSAR image based on Pauli decomposition.

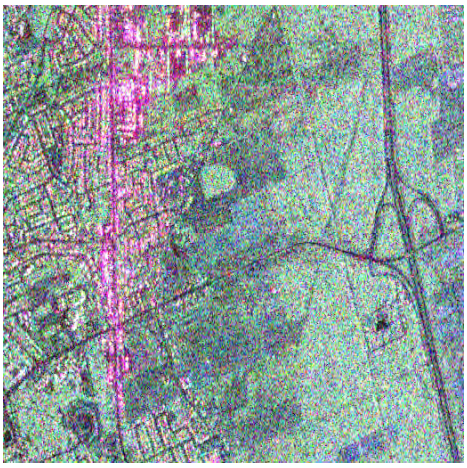


Figure 2. Original PolSAR image with Red = $|S_{HH}+S_{VV}|$, Green = $|S_{HH}-S_{VV}|$, Blue = $|S_{HV}|$

3.2 Experimental Results and Analysis

In the experiment, we use a 11×11 window with the const false alarming rate 0.005 to detect point target, and then apply the proposed method with a 7×7 window. Finally, filtered covariance matrix is obtained.

Covariance matrix can be transformed to coherent matrix of which the diagonal elements are the intensities of elements in polarimetric vector *k*. Therefore we also can obtain a colour image using Pauli decomposition. Figure 3 shows the filtered PolSAR image with Pauli decomposition colour composition where speckle is reduced greatly, and colour remains very similar to that in figure2.

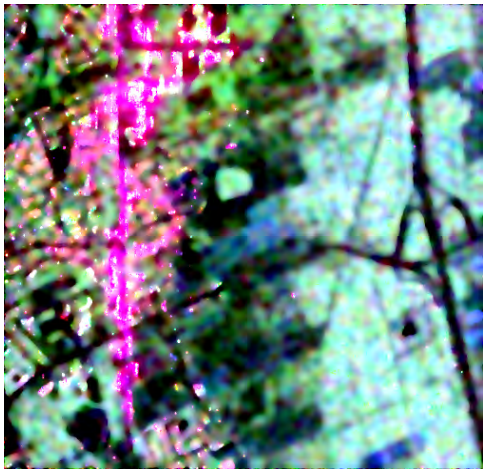


Figure 3. Filtered image using filtered proposed in this paper with Red = $|S_{HH}+S_{VV}|$, Green = $|S_{HH}-S_{VV}|$, Blue = $|S_{HV}|$

In order to test the validity of the proposed method, J.S. Lee filter is compared. Table 4 shows the mean, variance and the equivalent number of looks (ENL) for original, J.S. Lee filter and proposed filter in this paper. Both methods have similar performance for mean preservation and speckle reduction.

filter	original	J.S. Lee	proposed
mean	0.4928	0.4564	0.4876
variance	0.2880	0.0203	0.0262
ENL	0.84	10.27	9.06

Table 4. Mean and variance for intensity

The extent of preservation for polarimetric information is another important index for polarimetric filter. The ratio of different polarized channel is one of the most important polarimetric information indexes. In this study, in order to determine whether the proportion of each polarized channel is changed we use the percentage of each channel in SPAN to substitute the channel ratio. Table 5 shows the comparisons between original image and the two images filtered using J.S. Lee filter and the proposed filter. It shows that both methods can preserve the ratio between different polarizations. But the ratio for the proposed filter is closer to original ratio than that for J.S. Lee filter. It is demonstrated that polarimetric information is better preserved.

Ratio	E(HH/SPAN)	2×E(HV/SPAN)	VV/SPAN
original	30.86%	37.17%	31.97%
J.S. Lee	31.97%	37.59%	30.44%
proposed	31.23%	37.22%	31.54%

Table 5. Percentage for each channel in SPAN

4. CONCLUSIONS

A new filter framework for PolSAR imagery has been proposed in this paper. This algorithm is based on the multiplicative model of SPAN and polarimetric information for PolSAR data. SPAN and normalized covariance matrix are filtered respectively, and finally is combined to be the restored covariance matrix. Experimental results show that this method has the performance of speckle reduction similar to extended J.S. Lee filter and polarimetric information preservation better than extended J.S. Lee. Nevertheless, more studies to preserve edges, details in PolSAR filters propose in this paper are necessary.

REFERENCES

Frost, S.V., Stiles, A. J., Shanmugan, S. K., and Holtzman, C. J., 1982. A model for radar images and its application to adaptive digital filtering of multiplicative noise, *IEEE transations On pattern analysis and machine intelligenc*, 4(2), pp157-166.

Gigli, G., Lampropoulos, G. A., 2002. A new maximum likelihood generalized gamma CFAR detector, *IGARSS'02*, pp. 3399-3401.

Kuan, T.D., Sawchuk, A. A., Strand, C. T., and Chavel, P. , 1987. adaptive restoration of images with speckle, *IEEE transactions on ASSP*., 35(3), pp. 373-383.

Lee, J.S., Pottier, E., 2009. *Polarimetric radar imaging from basics to applications*. CRC Press, Boca Raton, pp. 53-84

Lee, J.S., 1980. Digital image enhancement and noise using local statistics, *IEEE transantions on pattern analysis and machine intelligence*, PAM1-2(2), pp. 165-168

Liu, G.Q., Huang, S.J., Toorre, A., and Rubertone, F., 1995. Optimal multi-look polarimetric speckle. *Proc. IGASS'95 Sym.*, pp. 664-666.

Novak, L.M., and Burl, M. C., 1990. Optimal speckle reduction in polarimetric SAR imagery, *IEEE trans. Aerosp. Electron. Syst.*, 26(2), pp. 293-305.

Vasile, G., Trouve, E., Lee, J.S., Vasile Buzuloiu, 2006. Intensity-driven adaptive neighbourhood technique for polarimetric and interferometric SAR parameters estimation, *IEEE trans. Geosci. Remote sensing*, 44(6), pp. 1609-1611.

ACKNOWLEDGEMENT

This study was supported by Project 2009AA12Z145 which was funded by the national high technology research and development program of China.

EVALUATION OF SPECTRAL AND TEXTURE FEATURES FOR OBJECT-BASED VEGETATION SPECIES CLASSIFICATION USING SUPPORT VECTOR MACHINES

Z. Li [a, b, *], R. Hayward [a], J. Zhang [a], H. Jin [a], R. Walker [b]

[a] Queensland University of Technology, George Street, Brisbane, QLD 4001 Australia
[b] Australian Research Centre for Aerospace Automation (ARCAA), Boronia Road, Brisbane, QLD 4009 Australia
(zhengrong.li, r.hayward, Jinglan.zhang, hang.jin, ra.walker)@qut.edu.au

KEY WORDS: Classification, Multispectral, Object, Feature, Vegetation

ABSTRACT:

The use of appropriate features to characterize an output class or object is critical for all classification problems. This paper evaluates the capability of several spectral and texture features for object-based vegetation classification at the species level using airborne high resolution multispectral imagery. Image-objects as the basic classification unit were generated through image segmentation. Statistical moments extracted from original spectral bands and vegetation index image are used as feature descriptors for image objects (i.e. tree crowns). Several state-of-art texture descriptors such as Gray-Level Co-Occurrence Matrix (GLCM), Local Binary Patterns (LBP) and its extensions are also extracted for comparison purpose. Support Vector Machine (SVM) is employed for classification in the object-feature space. The experimental results showed that incorporating spectral vegetation indices can improve the classification accuracy and obtained better results than in original spectral bands, and using moments of Ratio Vegetation Index obtained the highest average classification accuracy in our experiment. The experiments also indicate that the spectral moment features also outperform or can at least compare with the state-of-art texture descriptors in terms of classification accuracy.

1. INTRODUCTION

Interpretation of remotely sensed images has played an important role in vegetation mapping in the past decades, however the use of coarser spatial resolution satellite imagery have proven insufficient or inadequate for discriminating species-level vegetation in detailed vegetation studies (Yu et al., 2006). Airborne high spatial resolution imagery provides more information for detailed observation of vegetation. However, traditional classification algorithms based on single pixel analysis are often not capable of extracting the information we desire from high spatial resolution images. In recent years, object-based approaches become popular in high spatial resolution image classification, which has proven to be an alternative to the pixel-based image analysis and a number of publications suggest that better results can be expected (Blaschke, 2010). When applying object-based method to vegetation species classification, individual trees are expected to be segmented as image-objects and after that classification will be conducted in object-feature space.

The use of appropriate features to characterize an output class or object is fundamental for all classification problems. How to extract representative object-features in arbitrary-shaped regions is still an open issue for object-based image classification. Texture is a fundamental feature to describe image, but most texture descriptors are based on regular images or regular regions (e.g. small blocks) and do not consider the color information (Liu et al., 2006). Shape features are very significant features which are very close to human perception. However due to the inaccuracy of image segmentation and view angle variations, shape features are not widely used in natural image analysis. We believe that statistical measurement is a better way to summarize arbitrary-shaped image regions in object-based image classification. Color histograms are the most widely used statistical features in computer vision. They are often used for the illumination independent characterization of the color distribution of the pattern. However, color histograms do not exploit the spatial layout of the colors. A good way to include such lost information is to use statistical moments as features. Color moments improve the characterization the shape and color distribution of the pattern and have proven to be effective features under changing viewpoint and illumination (Moons, 2004).

Most previous feature extraction methods were conducted in original spectral bands (e.g. RGB color space), which were often fragile in visually complex environments. Incorporating domain knowledge might be a better way in real-world image analysis projects. From the literature review, the dominant method for interpreting vegetation biophysical properties from optical remote sensing data is through spectral vegetation indices. Plants have distinctive spectral signatures which is often modelled by combinations of reflectance measured in two or more spectral bands (Myneni et al., 1995). To our knowledge, little work has been done on utilizing vegetation indices as visual feature descriptors to combine multiple spectral bands for vegetation species classification from remote sensing images.

In this paper, we take the advantage of vegetation spectral properties and use spectral moment features for object-based vegetation species classification. To evaluate the usefulness of spectral moment features, the state-of-art texture features such as Gray-Level Co-Occurrence Matrix (GLCM) and Local Binary Patterns (LBP) are also extracted for comparison purpose. Different feature descriptors were compared by means of classification accuracy. A Support Vector Machines (SVM) classifier is employed for the classification in object-feature space. Multispectral images are collected in a power line

* Corresponding author.

corridor for vegetation management purposes and an extensive experiment on this dataset is conducted.

2. STUDY AREA AND DATA

The data used in this research were collected in rural Queensland Australia in October 2008 for research into vegetation management in power line corridors. The reason why we need species information of individual trees is that vegetation management in power line corridors is based on their potential risks to power lines. Some tree species are of particular interest and are generally categorized as undesirable and desirable species. For example, species with fast growing rates and that also have the potential to reach a mature height of more than four meters are defined as undesirable species. These undesirable species often pose high risks to power lines and therefore should be identified and removed.

The images were captured in a 1.5 kilometres corridor by a high resolution 3-CCD digital multi-spectral camera mounted on fixed wing aircraft. Figure 1 shows a mosaic of the test area generated from aerial images acquired from the trial. The four spectral bands of the camera are: NIR (800-966 nm), red (670-840 nm), green (540-640 nm), and blue (460-545 nm). The spatial resolution of the captured images is about 15 cm. The ground truth data of vegetation species were obtained from a field survey with domain experts' participation.

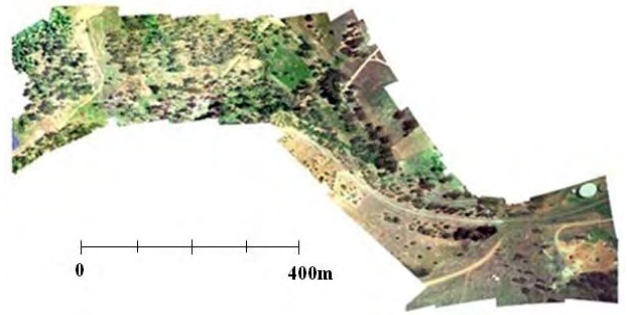


Figure 1. Experiment test site

It should be noted that classifying all types of species in power line corridors requires significantly more resources than are currently available, however, classifying species in a given test area as a proof of concept is possible. In this research, we focus on three dominant species in our test field: *Eucalyptus tereticornis*, *Eucalyptus melanophloia*, and *Corymbia tesselaris*. These three species Here we abbreviate the species names to *Euc_Ter*, *Euc_Mel* and *Cor_Tes*. According to the field survey, these three species account for over 80% of all the trees in the test corridor.

3. METHODS

In this research, object-based image analysis is adopted which consists of a three-stage processing: image segmentation, spectral and texture feature extraction, and supervised classification employing SVM.

3.1 Image Segmentation

Successful object-based image analysis results largely depend on the performance of image segmentation. Since we are going to classify the species among trees, tree crowns are the only image-objects of interest in our research. The aim of segmentation is, therefore, to detect and delineate all trees from images while eliminating other image regions. We have developed an automatic tree crown detection and delineation algorithm by utilizing spectral features (i.e. band ratio of near-infrared and red) in a pulse coupled neural network (PCNN) followed by post-processing using morphological reconstruction (Li et al., 2009). Since PCNN is capable to capture the proximity of image structure and texture, this method can automatically detect and delineate tree crowns from multi-spectral images and has been proved to be superior to some classic segmentation algorithms. Figure 2 shows an example of the segmentation results generated by our automatic segmentation algorithm.

Although the automatic segmentation is satisfied from visual assessment, decomposition of tree clusters is occasionally poor. Since the main aim of this research is evaluate the effectiveness of different feature descriptors for detailed vegetation species classification, manual segmentation is used to minimize the influence of inaccuracy in segmentation. The background is removed and each tree crown is labelled with a unique label to identify the tree which is paired against individual tree species obtained from field surveys. After segmentation, different feature descriptors are extracted from the segments (i.e. tree crowns) and used for training classifiers.

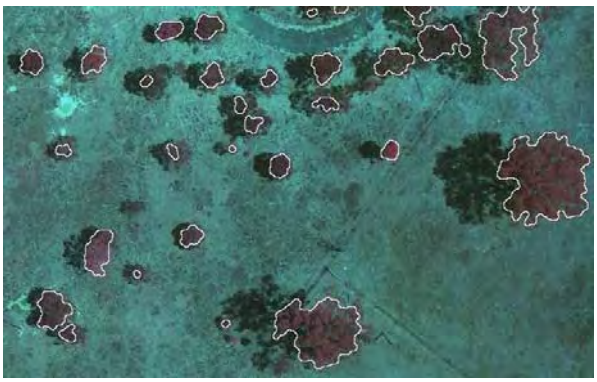


Figure 2. Example of automatic segmentation results

3.2 Spectral and Texture Feature Extraction

The object-based classification is substantially different from a per-pixel classification as it is done in object-feature space. Once the image-objects are segmented, both spectral and spatial attributes of each image-object (polygon) are extracted and used as input to a variety of classification algorithms for analysis. The basic approach to compute object-features from a multi-spectral image is to calculate separately the derivatives of the spectral channels. However, to generate features which could have high discriminative power among tree species is difficult as they all look green from visual spectrum. In addition, there can be large variations in lighting and viewing conditions for remotely sensed images, which may greatly affect the classification results if the feature descriptors used are not robust to these changes.

3.2.1 Spectral Moment Features: Color histograms are often used for the illumination independent characterization of the color distribution of the pattern. However, color histograms do not exploit the spatial layout of the colors. A good way to include such lost information is to use moments. Probability theory identifies that a probability distribution is uniquely characterized by its moments. Based on this idea, moment features have been proposed for color indexing (Stricker et al., 1995). However, moment features are mostly extracted from image as global features for image retrieval purpose, few work has been done on trying to represent image-object using moment features in object based image classification.

Since most information is concentrated on the low-order moments, only four central moments are considered as feature vectors in this research. They are defined as (Weinbach et al., 2007):

$$\mu = \frac{1}{\mathrm{N}}\sum_{i=1}^{N} s_i \tag{1}$$

$$\sigma = \sqrt{\frac{1}{\mathrm{N}}\sum_{i=1}^{N}(s_i-\mu)^2} \tag{2}$$

$$sk = \left(\frac{1}{N\times\sigma^3}\times\sum_{i=1}^{N}(s_i-\mu)^3\right)^{1/3} \tag{3}$$

$$ku = \left(\frac{1}{N\times\sigma^4}\times\sum_{i=1}^{N}(s_i-\mu)^4\right)^{1/4} \tag{4}$$

where in equations (1-4) N is total number of pixels within the image-object (i.e. tree crown), μ is the arithmetic mean, and σ represents the standard deviation, sk stands for the third moment skewness, and ku indicate the fourth moment kurtosis.

Plants have distinctive spectral properties. In the past decades, many spectral vegetation indices have been developed as measurements of relative abundance and activity of green vegetation. These vegetation indices are developed for purposes such as to estimate vegetation biophysical properties, to normalize or model external effects like viewing and sun angle variations and internal effects like background and soil variations (Jensen, 2000). Most of these vegetation indices are calculated from the near-infrared and red band of the spectrum. These vegetation indices have been successfully applied to measure biophysics of green vegetation. However, there has been very limited work on using these vegetation indices as feature descriptors for detailed vegetation species mapping, especially from the individual tree perspective.

In this paper, moments extracted from three widely used vegetation indices maps are evaluated: Ratio Vegetation Index (RVI) (Jordan, 1969), Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1973) and Perpendicular Vegetation Index (PVI) (Richardson et al., 1977). They are defined as:

$$RVI = \frac{\rho_{NIR}}{\rho_{red}} \tag{5}$$

$$NDVI = \frac{\rho_{NIR}-\rho_{red}}{\rho_{NIR}+\rho_{red}} \tag{6}$$

$$PVI = \frac{\rho_{NIR}-a\rho_{red}-b}{\sqrt{1+a^2}} \tag{7}$$

where ρ_{NIR} and ρ_{red} are the spectral reflectance of near-infrared and red band respectively. The parameters of PVI are set to be a=0.96916,b=0.084726 according to literature (Seo et al., 1998).

3.2.2 Texture Features: Texture contains important information in image classification, as it represents the content of many real-world images. Textures are characteristic intensity (or color) variations that typically originate from roughness of object surfaces (Davies, 2008). As a powerful source of information, texture features have been intensively studied in remote sensing image classification (Zhang et al., 2004, Franklin et al., 2000, Reulke et al., 2005, Samal et al., 2006). There are many different methods used to extract model texture from images. In this paper, we evaluated the widely used GLCM texture measures and state-of-art texture descriptor Local Binary Patterns (LBP) and its extensions: uniform LBP, rotation-invariant LBP, dominant local binary patterns (DLBP). In this section, an overview of these texture descriptors is given.

The image-objects generated from segmentation is arbitrary-shaped, however, texture measurements are usually extracted based on the texture property of pixels or small blocks within the rectangular shaped region. Therefore, in this paper, the arbitrary-shaped objects are extended to a rectangular area for texture extraction. This can be achieved by padding zero or mean value outside the object boundary, or obtaining the inner rectangle from the object. Zero padding introduces spurious high frequency components leading to degrading the performance of the texture feature, while the inner rectangle cannot usually represent the property of the entire object well. Mean-intensity padding has shown better performance than the other two approaches (Liu et al., 2006) and thus is adopted in this paper. Firstly, the minimum bounding rectangle is obtained from the image segment, and then the area which is outside of the segment and inside of the minimum bounding rectangle is padded using the mean value of pixels in the region.

Grey-level co-occurrence matrices (GLCM) have been successfully used for deriving texture measures from images. This technique uses a spatial co-occurrence matrix that computes the relationships of pixel values and uses these values to compute the second-order statistics (Haralick et al., 1973). The GLCM approach assumes that the texture information in an image is constrained in the overall or "average" spatial relationships between pixels of different grey level. In this paper, we use mean and standard deviation of four measures from the grey-level co-occurrence matrices: energy, entropy, contrast, and homogeneity.

LBP is first proposed by Ojala et al. to encode the pixel-wise information in the texture images (Ojala et al., 2002). The LBP method attempts to decompose the texture into small texture units and the texture features are defined by the distribution (histogram) of the LBP code calculated for each pixel in the region under analysis. Figure 3 gives an example of binary code in a 3×3 neighbourhood which generates 2^8 possible standard texture units. The LBP value for the centre pixel is calculated using the following equation:

$$LBP_{P,R} = \sum_{i=0}^{p-1} u(\mathrm{t}_i - t_c)\times 2^i \tag{8}$$

where P is the total number of neighbouring pixels, R is the radius used to form circularly symmetric set of neighbours. In this paper, we use $P=8, \mathrm{R}=1$.

Although LBP has proven to be a powerful texture descriptor, a number of extensions have been proposed to improve or supplement the classic LBP operators. We also evaluated several extensions to the conventional LBP operator including: uniform LBP, rotation-invariant LBP, and dominant LBP (DLBP) (Ojala et al., 2002, Liao et al., 2009). The uniform LBP is used to represent the most important microstructures, which contain at most two bitwise (0 to 1 or 1 to 0) transitions. The rotation-variant LBP is produced by circularly rotating the original the original LBP code until its minimum value is attained, making LBP code invariant with respect to rotation of the image domain. DLBP only considers the most frequently occurred patterns, and try to avoid the information loss caused by just considering the uniform LBP and the unreliability by considering all possible patterns.

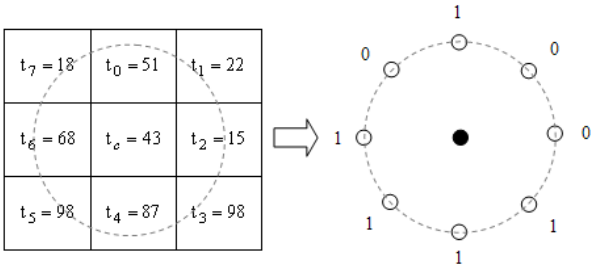


Figure 3. Example of binary code calculation in a 3×3 neighbourhood. The binary labels of the neighbouring pixels is obtained by applying a simple threshold operation with respect to the centre pixel t_c. $u(t_i - t_c)$ represents a step function, where $u(x) = 1$ when $x \geq 0$; else, $u(x) = 0$.

3.3 Supervised Classification Using SVM

In this research, the species distribution in the test area is known a priori through the field survey and thus supervised classification is adopted to evaluate the discriminative power of different features in vegetation species classification. From our field survey, a vegetation database has been generated by giving each tree in the test field a unique ID and recording several attributes of each tree (e.g. species name and values of all extracted object-features).

In our research, Support Vector Machines (SVMs) are employed as the classification methodology. SVM is an machine learning technology which has been successfully used in a variety of pattern recognition tasks and often outperforming other classification methodologies (e.g. Artificial Neural Networks) (Mills, 2008). The basic idea of SVM is to find an optimal decision function (a hyperplane) with the largest margin to separate the training data $\{x_1, x_2, ..., x_n\}$ with a label $y_i \in \{-1,+1\}$ into the positive (+1) or negative (-1) classes. The decision function is described as equation (9), and decision could be made according to that when $f(x) = 0$, x is classified as +1, otherwise, x is classified as -1. Figure 4 illustrates a simple linear separable case.

$$f(x) = w \cdot x + b \tag{9}$$

For data not linearly separable in the input space, SVM would map the data from the initial space to a (usually significantly higher dimensional) Euclidean space H by computation of inner-product kernels $K(x_i, x)$. After the mapping, the data, which is not linearly separable in the input space, become linearly separable in the H space. Thus, the SVM classifiers can be described as equation (10). Various classification methods are constructed by employing different kernel functions $K(x_i, x)$ (e.g., linear, polynomial, RBF, sigmoid, etc.). Radial basis function (RBF) is selected in this paper as it often suggested as the first choice since it has several advantages over other common kernel functions (Hsu et al., 2008): 1) unlike linear kernel, RBF nonlinearly maps samples into a high dimensional space, so it can handle the case when the relation between class labels and attributes is nonlinear; 2) RBF kernel has less hyperparameters than the polynomial kernel which make it less complex in model selection; 3)

$$f(x) = \sum_{i=1}^{n} y_i \alpha_i K(x_i, x) + b \tag{10}$$

$$\text{RBF kernel } K(x, y) = \exp(-\gamma \| x - y \|)^2 \tag{11}$$

where $0 \leq \alpha_i \leq C$ is the maximal margin hyperplane in the H space. When the maximal margin hyperplane is found, only those points that lie closest to the hyperplane have $0 \leq \alpha_i \leq C$, and these points are the *support vectors*.

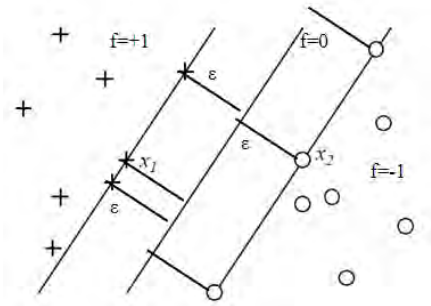
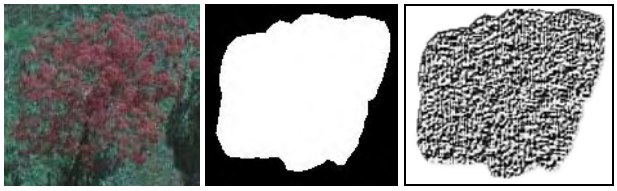


Figure 4. A linearly separable binary classification problem. The optimal hyperplane is with the maximum *margin* ε between the separating hyperplane and a hyperplane through the closest points of each of the two classes. These closest points are called the *support vectors* (x_1 and x_2 are examples of support vectors).

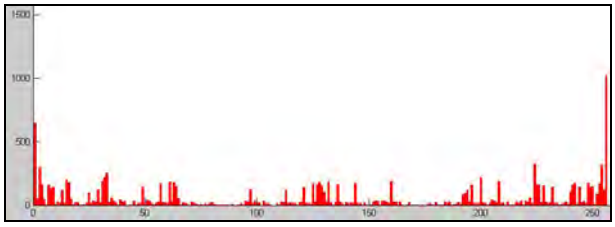
4. EXPERIMENT AND RESULT

4.1 Experiment Setup

The proposed spectral moment features are evaluated against the LBP and GLCM texture features on the multispectral data set discussed in section 2. The experiments are conducted in an open source SVM toolbox (SVMKM) (Rakotomamonjy et al., 2008). For the decision function of SVM, two parameters γ and C are specified using a grid search scheme. The 'one against one' strategy is employed for multi-class classification. The training samples include 75 trees with 25 for each species. Two testing datasets were used for evaluation with 60 samples in each dataset. Totally 10 region feature descriptors are extracted from the segments (polygons), of which LBPs and GLCM texture features are extracted from grey channel which is derived by averaging the four spectral bands. The LBP and its extensions are calculated in a 3×3 neighbourhood. All the feature descriptors are extracted from the regions of interest (segmented tree crowns). Figure 5 shows an example of LBP texture feature extraction from tree crowns. The extraction of other feature descriptors also follows the same procedure. Table 6 lists the evaluated features in the experiment, their abbreviations and feature dimensions.



(a) Original image (b) region of interest (c) LBP code image



(d) LBP histogram on the region of interest

Figure 5. Example of feature extraction from tree crowns

Feature	Abbr.	Dimension
Grey-level co-occurrence matrices	GLCM	8
Local Binary Pattern	LBP	256
Dominant Local Binary Patterns	DLBP	205
Uniform Local Binary Patterns	ULBP	59
Rotation-invariant Local Binary Patterns	ri_LBP	10
Spectral moments in RGB space	m_RGB	12
Spectral moments in CIR space	m_CIR	12
Spectral moments in RVI space	m_RVI	4
Spectral moments in NDVI space	m_NDVI	4
Spectral Moments in PVI space	m_PVI	4

Table 6. Evaluated features

4.2 Results and Discussion

The overall classification accuracy is obtained by comparing the classified data and the ground truth reference data. The overall accuracy is defined as:

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ samples} \quad (12)$$

Figure 6 compares the average classification accuracies in two datasets by using different feature descriptors. Classification accuracies in testing dataset1 using four central moments of RGB, CIR, RVI, NDVI and PVI spectral sub-space, LBP, uniform LBP, rotation-invariant LBP, DLBP and GLCM are 0.5, 0.533, 0.65, 0.567, 0.533, 0.45, 0.45, 0.467, 0.383, and 0.5 respectively. From the results we can see that the use of moments in spectral vegetation indices indicate higher classification accuracy than using original spectral bands and the state-of-art texture descriptors. Similar results were obtained in dataset2 with the average classification accuracy of 0.683, 0.717, 0.733, 0.717, 0.683, 0.733, 0.75, 0.617, 0.717, and 0.717 respectively for the 10 feature descriptors. From the experiment, we can see that incorporating spectral vegetation index in moment feature extraction improved the classification accuracy and the spectral moments in RVI showed the best performance.

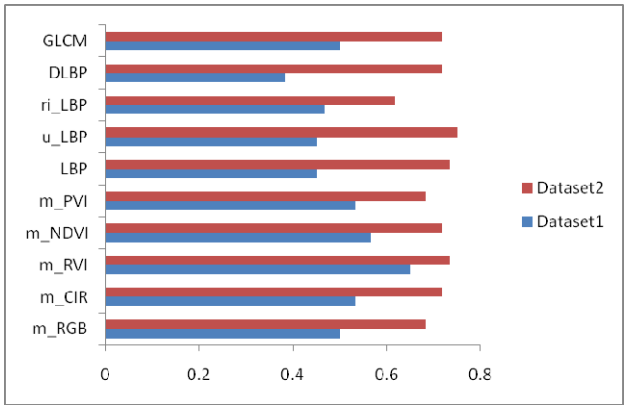


Figure 8. Evaluation of feature descriptors on two testing datasets. The horizontal axis indicates the average classification accuracy and the vertical axis compares different features descriptors.

Table 7 presents the average classification accuracies of the 10 features per category. It is noted that the evaluated features have different discriminative powers for different tree species. Therefore, it would be interesting to investigate whether the integration of multiple features will improve the classification result and how to select and fuse different features. A possible solution is to use feature subspace selection methods such as principal component analysis (Lu et al., 2007) and locally linear embedding (Roweis et al., 2001). These algorithms have been reported to be effective in reducing the dimensions of input space and achieving better performance which might be helpful when multiple features are used.

Trees can often show different appearances in different seasons and even the same tree species may vary due to the their health status. Nevertheless, from our experiment we can conclude that the spectral moment fetures derived from spectral index maps have the potential to improve the accuracy in detailed vegetation mapping tasks. Our future work is to fuse multiple spectral and texture features to further improve the classification accuracy.

Category	m_RGB	m_CIR	m_RVI	m_NDVI	m_PVI	LBP	ULBP	ri_LBP	DLBP	GLCM
Euc_Ter	0.725	0.675	0.65	0.8	0.65	0.375	0.375	0.325	0.35	0.4
Euc_Mel	0.6	0.6	0.675	0.625	0.575	0.65	0.475	0.425	0.45	0.475
Cor_Tes	0.45	0.6	0.75	0.5	0.6	0.75	0.95	0.875	0.85	0.95

Table 7. Overall classification accuracies of 10 features per category

5. CONCLUSION

This paper evaluates the capability of spectral moment and texture features for object-based vegetation species classification. Totally 10 spectral and texture feature descriptors were evaluated using SVM by means of classification accuracy. The experimental results showed that spectral moment features has the potential to improve the accuracy in individual tree species classification from high resolution multispectral images.

The use of spectral moment in RVI indicates the highest classification accuracy in our experiment.

REFERENCES

Blaschke, T., 2010. *Object-based image analysis for remote sensing*. ISPRS Journal of Photogrammetry & Remote Sensing 65(pp. 2-16

Davies, E. R., 2008. Introduction to Texture Analysis. Handbook of Texture Analysis Mirmehdi, M., Xie, X. and Suri, J. Imperial College Press. 1-31.

Franklin, S. E., Hall, R. J., Moskal, L. M.et al., 2000. *Incorporating texture into classification of forest species composition from airborne multispectral images.* International Journal of Remote Sensing 21(1), pp. 61-79

Haralick, R. M., Shanmugam, K. and Dinstein, I., 1973. *Textural features for image classification.* IEEE Transactions on Systems, Man, and Cybernetics 34(3), pp. 610-621

Hsu, C.-W., Chang, C.-C. and Lin, C.-J., 2008. *A practical guide to SVM classification (Technical Report).* Department of Computer Science, National Taiwan University.

Jensen, J. R., 2000. Remote sensing of vegetation. Remote Sensing of The Environment: An Earth Resource Perspective. Prentice Hall. 361-366.

Jordan, C. F., 1969. Derivation of leaf area index from quality of light on the forest floor. Ecology 50(pp. 663-666

Li, Z., Hayward, R., Zhang, J.et al., 2009. T*owards automatic tree crown detection and delineation in spectral feature space using PCNN and morphological reconstruction.* IEEE International Conference on Image Processing.

Liao, S., Law, M. W. K. and Chung, A. C. S., 2009. *Dominant Local Binary Patterns for Texture Classification*. IEEE Transactions on Image Processing 18(5), pp. 1107-1118

Liu, Y., Zhang, D., Lu, G.et al., 2006. Study on Texture Feature Extraction in Region-Based Image Retrieval System. International Conference on Multi-Media Modelling.

Lu, Y., Cohen, I., Zhou, X. S.et al., 2007. *Feature selection using principal feature analysis.* ACM International Conference on Multimedia

Mills, H., 2008. *Analysis of The Transferability of Support Vector Machines for Vegetation Classification.* The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. XXXVII 557-563.

Moons, F. M. T. T. L. V. G. T., 2004. Moment invariants for recognition under changing viewpoint and illumination. Computer Vision and Image Understanding 94(1-3), pp. 3-27

Myneni, R. B., Hall, F. G., Sellers, P. J.et al., 1995. *The interpretation of spectral vegetation Indexes.* IEEE Transactions on Geoscience and Remote Sensing 33(2), pp. 481-486

Ojala, T., Pietikainen, M. and Maenpaa, T., 2002. *Multiresolution grey-scale and rotation invariant texture classification with local binary patterns*. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(7), pp. 971-987

Rakotomamonjy, A. and Canu, S., 2008. SVM and Kernel Methods Matlab Toolbox. http://asi.insa-rouen.fr/enseignants/~arakotom/toolbox/index.html.

Reulke, R. and Haala, N., 2005. *Tree species recognition with fuzzy texture parameters*. Lecture Notes in Computer Science. Springer Berlin 3322: 607-620.

Richardson, A. J. and Wiegand, C. L., 1977. D*istinguishing vegetation from soil background information.* Photogrammetric Engineering & Remote Sensing 43 pp. 1541-1552

Rouse, J. W., Haas, R. H., Schell, J. A.et al., 1973. *Monitoring vegetation systems in the great plains with ERTS.* the 3rd Earth Resources Technology Satellite-1 Symposium. Washington, D.C., USA. NASA SP-351 pp. 309-317

Roweis, S. T. and Saul, L. K., 2001. Nonlinear dimensionalityr reduction by locally linear embedding. Science 290(22), pp. 2323-2326

Samal, A., Brandle, J. R. and Zhang, D., 2006. *Texture as the basis for individual tree identification.* Information Sciences 176 pp. 565-576

Seo, D.-J., Park, C.-H., Park, J.-H.et al., 1998. A *search for the optimum combination of spatial resolution andvegetation indices*. IEEE Geoscience and Remote Sensing Symposium pp. 1729 - 1731

Stricker, M. and Orengo, M., 1995. Similarity of color images. SPIE Conference on Storage and Retrieval for Image and Video Databases. 2420: 381-392.

Weinbach, R. W. and Richard M. Grinnell, J., 2007. Statistics for social workers. Pearson/Allyn & Bacon. 60-62.

Yu, Q., Gong, P., Clinton, N.et al., 2006. *Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery*. Photogrammetric Engineering & Remote Sensing 72(7), pp. 799-811

Zhang, C., Franklin, S. E. and Wulder, M. A., 2004. *Geostatistical and texture analysis of airborne-acquired images used in forest classification.* International Journal of Remote Sensing 25(4), pp. 859-865

ACKNOWLEDGEMENT

This work was conducted within the CRC for Spatial Information, established and supported under the Australian Government's Cooperative Research Centers Programme, and in conjunction with the Australian Research Centre for Aerospace Automation (ARCAA). The authors would like to thank Bred Jeffers from Greening Australia for assisting the field survey.

EMPIRICAL COMPARISON OF MACHINE LEARNING ALGORITHMS FOR IMAGE TEXTURE CLASSIFICATION WITH APPLICATION TO VEGETATION MANAGEMENT IN POWER LINE CORRIDORS

Zhengrong Li [a, b]*, Yuee Liu [a], Ross Hayward [a], Rodney Walker [b]

[a] School of IT, Queensland University of Technology, George Street, Brisbane, QLD 4001 Australia
[b] Australian Research Centre for Aerospace Automation (ARCAA), George Street, Brisbane, QLD 4001 Australia
(zhengrong.li, yuee.liu, r.hayward, ra.walker)@qut.edu.au

Commission VII

KEY WORDS: Classification, Texture Feature, Machine Learning, Object-based Image Analysis, Vegetation

ABSTRACT:

This paper reports on the empirical comparison of seven machine learning algorithms in texture classification with application to vegetation management in power line corridors. Aiming at classifying tree species in power line corridors, object-based method is employed. Individual tree crowns are segmented as the basic classification units and three classic texture features are extracted as the input to the classification algorithms. Several widely used performance metrics are used to evaluate the classification algorithms. The experimental results demonstrate that the classification performance depends on the performance matrix, the characteristics of datasets and the feature used.

1. INTRODUCTION

Vegetation management activities in power line corridors including tree trimming and vegetation control is a significant cost component of the maintenance of electrical infrastructure. Currently, most vegetation management programs for distribution systems are calendar-based ground patrol (Russell et al., 2007). Unfortunately, calendar-based tree trimming cycles are expensive. It also results in some zones being trimmed more frequently than needed and others not cut often enough. Moreover, it is seldom practicable to measure all the plants around power line corridor by field methods. Satellites and aerial vehicles can pass over more regularly and automatically than the ground patrol. Therefore, remotely sensed data have great potential in assisting vegetation management in power line corridors (Li et al., 2008). Remote sensing image classification is one of the key tasks for extracting useful information to assist power line corridor monitoring.

Texture contains important information for image classification, as it represents the content of many real-world images. Texture feature extraction and classification have been intensively studied for interpreting vegetation properties from remote sensing imagery (Franklin et al., 2000, Coburn and Roberts, 2004). Selection of appropriate texture measurements and classification algorithm are two critical steps in a texture classification problem. However, most previous research focused on how to representing texture in an image, few research verified the discriminatory power of different classification algorithms using these texture features. Lu and Weng reviewed a number of image classification techniques for improving classification performance and suggested that the use of multiple features and selection of suitable classification method are especially significant for improving the classification accuracy. However, no empirical comparison and quantitative results have been presented. It would be interesting to investigate which one have more impact on the classification results, the features or the classifiers?

Machine learning techniques are now widely used in remote sensing classification. A machine learning algorithm is one that can learn from experience (observed examples) with respect to some class of tasks and a performance measure (Mitchell, 1997). Different performance metrics are often used and it is possible for one learning method to perform well on one metric, but be suboptimal on other metrics. For example, SVMs are designed to optimize accuracy, whereas neural networks typically optimize squared error or cross entropy (Caruana and Niculescu-Mizil, 2004). Moreover, in many applications Accuracy are used as the only measure to assess the performance of the built classifier. However, there are many other evaluation methods such as Precision/Recall and ROC analysis. We need to understand the advantage and disadvantage of these measures before using them for evaluation. Sometimes we may need to find tradeoffs on these methods and try to select a model that best suit the problem.

The motivation behind this paper is to develop a better understanding of the machine learning process in object-based image classification, to evaluate the performance of different machine learning algorithms in a specific texture classification application, and to compare the results not only in terms of their classification accuracy but also the benefit and cost and some other properties such as computational cost.

2. METHODOLOGY

2.1 An overview of object-based image classification

Since remote sensing images consist of rows and columns of pixels, conventional land-cover mapping has been based on a per-pixel basis (Mas et al., 2006). Unfortunately, classification

* Corresponding author.

algorithms based on single pixel analysis often are not capable of extracting the information we desire from high spatial resolution images. For example, the spectral complexity of urban land-cover materials results in specific limitations using per-pixel analysis for the separation of human-made materials such as roads and roofs and natural materials such as vegetation, soil, and water (R.Jensen, 2005). We need information about the characteristics of a single pixel but those of the surrounding pixels so that we can identify areas (or segments) of pixels that are homogeneous. Object-based approaches become popular in high spatial resolution remote sensing image classification, which has proven to be an alternative to the pixel-based image analysis and a large number of publications suggest that better results can be expected (Blaschke, 2010).

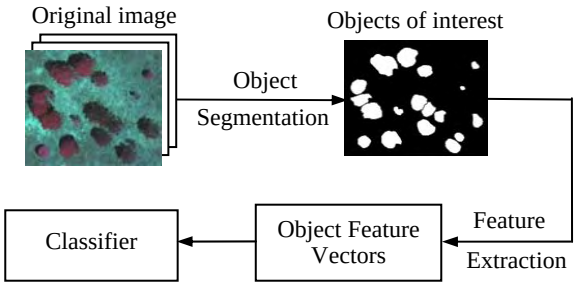


Figure 1 Framework of object-based image classification

A typical object-based image classification consists of a three-stage processing: image segmentation, object feature extraction, and pattern classification. Successful object-based image analysis results largely depend on the performance of image segmentation. Since we are going to classify the species among trees, tree crowns are the only image-objects of interest in our research. The aim of segmentation is, therefore, to detect and delineate all trees from images while eliminating other image regions. We have developed an automatic tree crown detection and delineation algorithm by utilizing spectral features in a pulse coupled neural network followed by post-processing using morphological reconstruction (Li et al., 2009). Although the automatic segmentation is satisfied from visual assessment, decomposition of tree clusters is occasionally poor. Since the main aim of this research is evaluate different machine classifiers, manual segmentation is used to minimize the influence of under-segmentation and over-segmentation. The background is removed and each tree crown is labelled with a unique label to identify the tree which is paired against individual tree species obtained from field surveys.

2.2 Texture Feature Extraction Methods

Texture patterns are defined as the characteristic intensity variations that typically originate from roughness of object surfaces (Davies, 2009). According to a recent review texture feature extraction methods can be divided into three categories: statistical, structural and signal processing based approaches (Xie and Mirmehdi, 2009). In this paper, three widely used texture features are extracted from the segments (polygons) and then input to the classifiers: GLCM, Gabor wavelet features, and Uniform LBP. In this paper, all three texture features are extracted from grey channel which is derived by averaging the four spectral bands of the original image.

GLCM: Grey-level co-occurrence matrices (GLCM) have been successfully used for deriving texture measures from images. This technique uses a spatial co-occurrence matrix that computes the relationships of pixel values and uses these values to compute the second-order statistics (Kubo et al., 2003). In this paper, we use mean and standard deviation of four measures from the grey-level co-occurrence matrices: energy, entropy, contrast, and homogeneity. The GLCM feature vector has 8 dimensions.

Gabor Wavelet Features: 24 Gabor wavelet filters are employed with center frequencies [0.05, 0.4], 4 scaling factors, and 6 orientations at angles of 0 and 180 degrees to achieve optimal coverage in the Fourier domain. The mean and standard deviation of magnitude of each filtered image region are used as feature components. The feature vector has 48 dimensions.

ULBP: Local Binary Pattern (LBP) is first proposed by Ojala et al. to encode the pixel-wise information in the texture images (Ojala et al., 2002). The LBP value for the centre pixel is calculated using the following equation:

$$LBP_{P,R} = \sum_{i=0}^{p-1} u(u_i - u_c) \times 2^i \qquad (1)$$

where P is the total number of neighbouring pixels, R is the radius used to form circularly symmetric set of neighbours. In our experiment, we use the uniform LBP (ULBP) contains at most two bitwise (0 to 1 or 1 to 0) transitions The occurrence histograms of the ULBP are computed using $P = 8, 16, 24$, with $R = 1, 2, 3$ respectively, which is claimed to have the best performance of the local binary patterns in the experiments conducted by Ojala *et al.* (Ojala et al., 2002). The features are obtained by combining the three sets of features together.

2.3 Machine Learning Algorithms

During the past decades, a variety of machine learning algorithms have been proposed for classification tasks. Although the potential advantages and disadvantage of these techniques have been addressed in many published work, most of them are from the theoretical view under some assumption about data distribution, characteristics of the classification task, signal-to-noise-ratio, etc. In reality, these assumptions are often hard to be verified. Therefore, a practical solution for selecting an appropriate model for a given classification task is to experimentally compare these algorithms. In this paper, we compared seven widely used machine classifiers which are implemented in DTREG (Sherrod, 2009): K-Means Clustering, Linear Discriminant Analysis (LDA), Radial Basis Function Networks (RBFN), Multilayer Perceptron Neural Networks (MLPNN), Support Vector Machines (SVM), Single Decision Tree (SDT), and Decision Tree Forest (DTF). Only a brief introduction of these algorithms is presented in this section, and may safely be skipped by readers since they are all well known techniques.

K-Means Clustering (KM): K-Means is a classic unsupervised clustering technique. When used for supervised classification, the model is built by minimizing the classification error (distances between the predicted cluster and the actual cluster membership. In DTREG, the training is done by searching the optimal number of clusters and each category may have several corresponding clusters.

Linear Discriminant Analysis (LDA) The basic idea of Linear Discriminant Analysis (LDA) is to find the linear combination of features (“linear transformation”) which best separate desired classes.

Multilayer Perceptron Neural Networks (MLP): Neural networks are predictive models loosely based on the action of biological neurons. Artificial neural network usually refers to multilayer percetron neural network which is typically full-

connected, three layers, feed forward, perceptron neural network.

Radial Basis Function Networks (RBFN): The basic idea of RBFN is that a predicted target value of an item is likely to be about the same as other items that have close values of the predictor variables. A RBFN typically has three layers: an input layer for each predictor variable, a hidden layer that uses Gaussian function as radial activation function and an output layer that implements weighted sum of hidden layer outputs.

Support Vector Machines (SVMs): The basic idea of SVM is to find an optimal decision function (a hyperplane) that separates clusters of vector in such a way that cases with one category of the target variable are on one side of the plane and cases with the other category are on the other size of the plane. The vectors near the hyperplane are the support vectors.

Single Decision Tree (SDT): Decision tree is a binary tree structure whose internal nodes correspond to input patterns and whose leaf nodes are categories of patterns. The tree can be induced by iteratively splitting the dataset into subsets based on classes attributes. The decision tree assigns a pattern category to an input pattern by filtering the pattern from the root to the leaf in the tree.

Decision Tree Forests (DTF): It is also known as Random Forests, which is an ensemble of tree-type classifiers. A decision tree forest grows a number of independent trees in parallel, and they do not interact until after all of them have been built. For classification, each tree in the DTF casts a unit vote for the most popular class at input, while the output of the classifier is determined by a majority vote of the trees.

2.4 Performance Metrics

Given a certain application, more than one method is applicable. This motivates evaluating the performance of these classification methods empirically in a specific application. That is, given several classification algorithms, how can we say one has less error than the others for a given application? Having selected a classification algorithm to train a classifier, can we tell an expected error rate with enough confidence that later on when it is used in a new dataset?

In this section, we consider several most commonly used metrics for evaluating different classification algorithms: overall accuracy, precision/recall, F-measure, ROC analysis, and computational cost. All of these measures are based on the definition of a confusion matrix. An example of confusion matrix for binary classification is described in Table 1. To help the definition that follows, we define the following symbols: TP: True Positive count; FN: False Negative count; FP: False Positive count; TN: True Negative count.

The overall accuracy is the simplest and most intuitive evaluation measure for classifiers. It is defined as

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ samples} = \frac{TP+TN}{P+N}$$

It is worth noting that the overall accuracy does not distinguish between types of errors the classifier makes (i.e. False Positive versus False Negative) (Japkowicz, 2006). For example, two classifiers may obtain the same accuracy but they may behave quite differently on each category. If one classifier obtains 100% accuracy on one category but only 41% on the other category, while another classifier generate 70% for each category, it is hard to claim that the first classifier is better. Therefore, overall accuracy may not be use blindly as the evaluation method for classifiers on a dataset. Precision and Recall can avoid the problem encountered by Accuracy. Precision can be seen as a measure of exactness or fidelity, whereas Recall is a measure of completeness. Their definitions are: $\Pr ecision = TP/(TP+FP)$, $\mathrm{Re}call = TP/P$. Usually, Precision and Recall scores are discussed jointly and a single measure can be derived by combing both measures (e.g. F-measure). F-measure is the weighted harmonic mean of precision and recall. In this paper, we use the F_1 measure in which the precision and Recall are evenly weighted. It is defined as:

$$F_1 = 2 \cdot \frac{\Pr ecision \cdot \mathrm{Re}call}{\Pr ecision + \mathrm{Re}call}$$

The goal of Precision/Recall space is to be in the upper-right-hand corner, which means that the higher value of F_1 measure, the better classifier's performance.

Table 1 A confusion matrix

Predicted Category	Actual Category	
	Positive	Negative
Positive	TP	FP
Negative	FN	TN
	P=TP+FN	N=FP+TN

Precision and Recall do not judge how well a classifier decides that a negative example is, indeed, negative. Receiver Operating Characteristic (ROC) analysis can solve both the problems of Accuracy and Precision/Recall. ROC analysis plots the False Positive Rate (FPR) on the x-axis of a graph and True Positive Rate (TPR) on the y-axis. TPR is equal to Recall and FPR is defined as $FPR = FP/N$. An ROC graph depicts relative trade-offs between true positive (benefits) and false positive (costs), and the goal in ROC space is to be in the upper-left-hand corner (Davis and Goadrich, 2006). The (0,1) point of the ROC space is also called a perfect classification. The diagonal line from the left bottom to the right top corner is also called the random guess line, which can be used to judge the whether it is good or bad classification. Points above the random guess line indicate good classification results, while points below the line are considered as bad classification results. In this paper, we calculate the distance of the each point and the (0,1) point and rank it. The shorter the distance, the better the classification is.

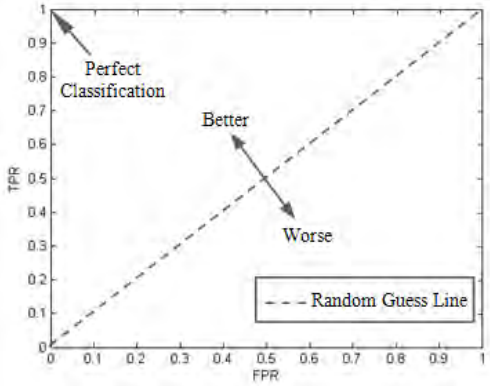


Figure 2 Illustration classifier evaluation in ROC space

Computational costs of the classification algorithms also need to be considered in a real-world problem. Although in most remote sensing image classification tasks real-time processing is not required, it is certainly not unnecessary to choose a computational efficient classification algorithm. In this paper, we compare the computation cost of different machine learning

algorithms by recording the analysis time in both training and testing stages.

3. EXPERIMENT AND DISCUSSION

3.1 Data Collection

The experiment dataset used in this research were collected in rural Queensland Australia in October 2008 for research into vegetation management in power line corridors. The reason why we need species information of individual trees is that vegetation management in power line corridors is based on their potential risks to power lines. Some tree species are of particular interest and are generally categorized into undesirable and desirable species. For example, species with fast growing rates and that also have the potential to reach a mature height of more than four meters are defined as undesirable species. These undesirable species often pose high risks to power lines and therefore should be identified and removed. The images were captured in a 1.8 kilometres corridor by a high resolution 3-CCD digital multi-spectral camera mounted on fixed wing aircraft. Figure 3 shows a mosaic of the test area generated from aerial images acquired from the trial. The spatial resolution of the captured images is about 15cm. The ground truth data of vegetation species were obtained from a field survey with domain experts' participation.

It should be noted that classifying all types of species in power line corridors requires significantly more resources than are currently available, however, classifying species in a given test area as a proof of concept is possible. In this research, we focus on three dominant species in our test field: *Eucalyptus tereticornis*, *Eucalyptus melanophloia*, and *Corymbia tesselaris*. We abbreviate the species names to *Euc-Ter*, *Euc-Mel* and *Cor-Tes*. Through field survey with botanist's participation, 121 trees were selected and labelled for the experiment with 64 Euc-Ter, 30 Euc-Mel and 27 Cor-Tes. The criterion is that tree crowns are big enough so that they can be visually identified from the aerial images. Visual classification of these species often uses features such as leaf shape and bark type which are not available from the data used. However, texture analysis can be very useful to identify these species from digital imagery.

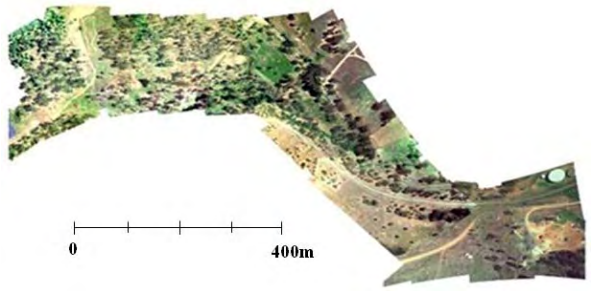


Figure 3 Experiment test site

3.2 Results and Discussion

To evaluate the performance of different machine learning algorithms in texture classification, we use the implementation of these algorithms in DTREG. For all classifiers the default setting of DTREG is used. V-fold cross validation technique is employed in the experiment, and 10 folders were selected for the cross validation. The dataset is partitioned into 10 groups, which is done using stratification methods so that the distributions of categories of the target variable are approximately the same in the partitioned groups. 9 of the 10 partitions are collected into a pseudo-learning dataset and a classification model is built using this pseudo-learning dataset. The rest 10% (1 out of 10 partitions) of the data that was held back and used for testing the built model and the classification error for that data is computed. After that, a different set of 9 partitions is collected for training and the rest 10% is used for testing. This process is repeated 10 times, so that every sample has been used for both training and testing. The classification accuracies of the 10 testing datasets are averaged to obtain the overall classification accuracy.

Table 2 summarizes the overall classification accuracy of each machine classifier on the three feature vectors respectively. As is shown in the experimental results, of the seven methods investigated in this paper, the left three (KM, LDA and RBFN) show relatively low overall classification accuracy, whereas the MLP SVM classifiers generate higher accuracy on all three features. It is also noted that the SDT and DTF methods also give relatively good results when using Gabor and ULBP features, however, the classification accuracy drop off considerably when using GLCM features.

We also compare the average F_1 measure of three categories from different classification algorithms (Figure 4). As discussed in the previous section, a higher value of F_1 measure indicates a better classifier. From the figure it is clear that MLP and SVM generally perform well for all three features, while the performance of other classifiers largely depends on the data used. For example, RBFN obtains reasonable result for Gabor and ULBP features but generates terrible result when using GLCM feature.

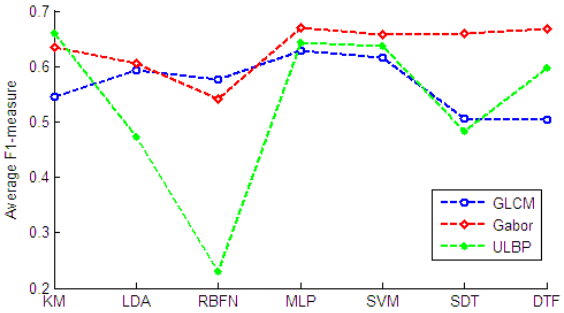


Figure 4 Average F1 measure of different classifiers

Figure 5 presents the analysis results of different classification algorithms for three texture features in ROC space. The plots of different algorithms use different markers specifiers, and within which the three categories are shown as different colours. Different from the analysis results of using overall accuracy and F_1 measure (Precision/Recall), ROC space provide more details of the classifier performance. As we can see from the figures, the performance of the classifier depends on the category and the feature used. By calculating the distance of points to the upper-left-corner point (point (0,1) in ROC space), the performance of the classifiers is ranked. From the experimental results, most classifiers perform the best for classifying *Cor-Tes*. The MLP classifier with ULBP features, the KM classifier with GLCM features, and the KM classifier with ULBP features obtained the best performance for classifying *Euc-Ter*, *Euc-Mel* and *Cor-Tes* respectively. The analysis result from ROC space is different from that derived from overall Accuracy and average F_1 measure, where SVM and MLP are supposed to be superior to other classifiers.

Table 2 Comparison of the overall classification accuracies

Features \ Classifiers	KM	LDA	RBFN	MLP	SVM	SDT	DTF
GLCM	55.37	64.46	62.81	69.42	69.42	58.68	56.20
Gabor	65.29	62.81	57.02	71.90	71.07	71.90	71.07
ULBP	69.42	50.41	52.89	72.73	71.07	66.12	71.07

Table 3 Comparison of the computational costs (in seconds)

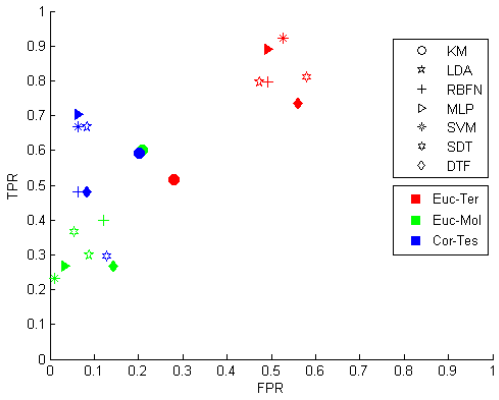
Features \ Classifiers	KM	LDA	RBFN	MLP	SVM	SDT	DTF
GLCM	2.64	0.23	43.53	2.72	22.89	0.3	0.55
Gabor	44.06	0.47	139.14	5.81	15.97	0.56	1.13
ULBP	385.97	7.41	113.19	136.41	230.93	2.53	2.31

Table 3 compares the computational cost of each machine classifier on the three feature vectors respectively. The analysis time is recorded by DTREG software under a desktop PC configuration of core duo 2.66GHz CUP and 2GB memory. From the results, we can see that the analysis time varies a lot for each machine classifiers and feature vectors. Overall, LDA, SDT and DTF are very computational efficient, whereas RBFN, MLP and SVM are computational much more intensive. It should also be mentioned that with the dimensions of feature vectors increase, the computational cost increase considerably (The dimensions of GLCM, Gabor and ULBP are 8, 48 and 607 respectively). For example, the analysis time of KM algorithm increase considerably when using ULBP feature.

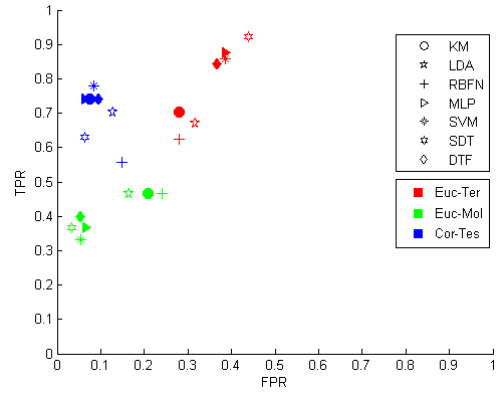
From the evaluation results, it is noticed that: 1) The selection of an approriate performance metrix is critical to evaluate the discriminatory power of different classifiers. Simply choose accuracy as the only measure often cause some misleading evaluation results. ROC analysis provide more details about the benefit and cost of a classifier. 2) The classification performance not only depends on the discriminatory power of classifiers but also the characteristics of datasets and the feature(s) selected. The evaluation results suggest to select approprate feature and classification algorithm for different categories. For example, to classify *Euc-Ter* the MLP classifier and ULBP feature are suggested. 3) Choosing a 'best model' is a complex issue and need to consider many factors such as the tradeoff between discriminatory power and computational cost. 4) Overall, the classification accuracies of all classifiers and texture features are not as good as expected. Trees can often show different apperances in different seasons and even the same tree species may vary due to the their health status. Nevertheless, using texture feature and machine learning techniques has shown the potential in analyzing vegetation in power line corridors by means of digital remote sensing imagery.

4. CONCLUSION

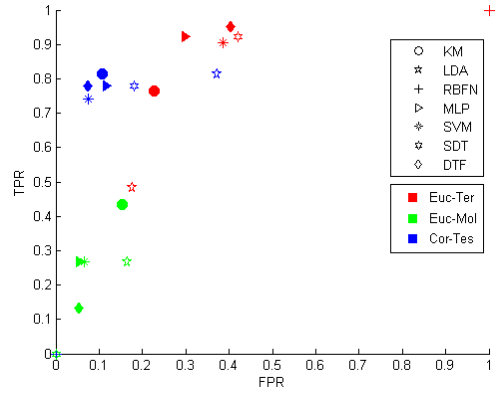
This paper evaluates the capability of seven machine learning algorithms and 3 texture features by means of classifying vegetation species in a power line corridor using high resolution aerial imagery. Object-based method is employed that local texture features are extracted from image-objects (i.e. tree crowns) and the classification is done in object feature space. Several performance matrixes are used to evaluate the performance of classifiers. The experimental results showed that the classification performance depends on the performance matrix, the characteristics of datasets and the feature(s) used.



(a) GLCM feature



(b) Gabor feature



(c) ULBP feature

Figure 5 The analysis results in ROC Space

ACKNOWLEDGEMENT

This work was conducted within the CRC for Spatial Information, established and supported under the Australian Government's Cooperative Research Centers Programme, and in conjunction with the Australian Research Centre for Aerospace Automation (ARCAA). The authors would like to thank Bred Jeffers from Greening Australia for assisting the field survey. The Authors would also like to acknowledge Ray Duplock for his support in high performance computing resources in QUT.

REFERENCES

Blaschke, T., 2010. Object-based image analysis for remote sensing. ISPRS Journal of Photogrammetry & Remote Sensing 65(1), pp. 2-16

Caruana, R. and Niculescu-Mizil, A., 2004. Data mining in metric space: an empirical analysis of supervised learning performance criteria. International Conference on Knowledge Discovery and Data Mining. Seattle, USA pp.

Coburn, C. A. and Roberts, A. C. B., 2004. A multiscale texture analysis procedure for improved forest stand classification. International Journal of Remote Sensing 25(20), pp. 4287-4308

Davies, E. R., 2009. Introduction to Texture Analysis. Handbook of Texture Analysis Mirmehdi, M., Xie, X. and Suri, J. Imperial College Press. 1-31.

Davis, J. and Goadrich, M., 2006. The relationship between Precision-Recall and ROC curve. the 23rd International Conference on Machine Learning. Pittsburgh pp. 233-240

Franklin, S. E., Hall, R. J., Moskal, L. M.et al., 2000. Incorporating texture into classification of forest species composition from airborne multispectral images. International Journal of Remote Sensing 21(1), pp. 61-79

Japkowicz, N., 2006. Why question machine learning evaluation method? AAAI workshop on Evaluation Methods for Machine Learning AAAI Press. 6-11.

Kubo, M., Kanda, F. and Muramoto, k., 2003. Texture feature extraction of tree using co-occurrence matrix from aerial images. SCIE Annual Conference. The Society of Instrument and Control Engineers.

Li, Z., Hayward, R., Zhang, J.et al., 2008. Individual tree crown delineation techniques for vegetation management in power line corridor. Digital Image Computing: Techniques and Applications (DICTA). Canberra pp. 148-154

Li, Z., Hayward, R., Zhang, J.et al., 2009. Towards automatic tree crown detection and delineation in spectral feature space using PCNN and morphological reconstruction. IEEE International Conference on Image Processing.

Mas, G. Y. J. F., Maathuis, B. H. P., Zhang, X.et al., 2006. Comparison of pixel-based and object-oriented image classification approaches - a case study in a coal fire area, Wuda, Inner Mongolia, China. International Journal of Remote Sensing 27(18), pp. 4039-4055

Mitchell, T., 1997. Machine Learning. McGraw Hill.

Ojala, T., Pietikainen, M. and Maenpaa, T., 2002. Multiresolution grey-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(7), pp. 971-987

R.Jensen, J., 2005. Classification based on object-oriented image segmentation. Introductory Digital Image Processing: A Remote Sensing Perspective (Third Edition). Pearson Education. 393-398.

Russell, B. D., Benner, C. L., Wischkaemper, J.et al., 2007. Reliability based vegetation management through intelligent system monitoring. Power Systems Engineering Research Center. Texas A&M University.

Sherrod, P. H., 2009. DTREG predictive modeling software. Users Manual. www.dtreg.com\dtreg.pdf.

Xie, X. and Mirmehdi, M., 2009. A galaxy of texture features. Handbook of Texture Analysis Mirmehdi, M., Xie, X. and Suri, J. Imperial College Press. 375-406.

ASSESSMENT OF THE MULTIPLE ENDMEMBER SPECTRAL MIXTURE ANALYSIS (MESMA) MODEL APPLIED TO THE HYPERION/EO-1 HYPERSPECTRAL DATA OF THE COASTAL PLAIN OF RIO GRANDE DO SUL, BRAZIL

R. M. Linn [a], S. B. A. Rolim [a,*], L. S. Galvão [b]

[a] Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil – (rodrigo.linn, silvia.rolim)@ufrgs.br
[b] Divisão de Sensoriamento Remoto, Instituto Nacional de Pesquisas Espaciais, São José dos Campos, Brazil - lenio@dsr.inpe.br

KEY WORDS: Land Cover, Radiometry, Classification, Processing, Pattern, Hyper spectral

ABSTRACT:

The objective of this work was to evaluate the potential use of the Multiple Endmember Spectral Mixture Analysis (MESMA) when applied to EO-1 Hyperion hyperspectral data to discriminate land covers in the southern state of Rio Grande do Sul, Brazil. The methodology involved: (a) pre-processing and atmospheric correction of Hyperion data; (b) sequential use of the Minimum Noise Fraction (MNF), Pixel Purity Index (PPI) and n-Dimensional Visualizer techniques in the 454-2334 nm range for the initial selection of a general group of endmember candidates (first spectral library) and of another group of pixels to be used for model validation; (c) use of the Visualization and Image Processing for Environmental Research Tools (VIPER Tools) to perform the final selection of endmembers based on the first spectral library and to obtain MESMA models; and (d) evaluation of resultant fraction images and root mean square error (RMSE) values to determine the optimal number of components of the MESMA model. Results showed that a four-endmember MESMA model (soil = dunes and dry fields; green vegetation = pinus, eucalyptus and grasslands; water = without sediments, with sediments, and with chlorophyll; and shade) adequately described the diversity of the scene components, including that of materials within the same class (e.g., pinus and eucalyptus) and produced the largest fractions and the lowest RMSE values on a per-pixel basis. Results demonstrated the potential use of the MESMA with EO-1 Hyperion hyperspectral data to discriminate land covers in the coastal plains of Rio Grande do Sul, even considering the low signal-to-noise ratio of the instrument, especially in the shortwave infrared range.

1. INTRODUCTION

Numerous techniques for the classification of orbital images have been used in the extraction of land surface data (land covers and land use, for instance). Multispectral sensors such as the Enhanced Thematic Mapper (ETM+, Landsat 7) with 8 spectral bands are generally used in such studies. However, these sensors do not allow for a more detailed study of the spectral behavior of the targets.

On November 21st, 2000, the Hyperion sensor was launched aboard the National Aeronautics and Space Administration's (NASA) Earth Observing-1 satellite (EO-1), becoming the first hyperspectral sensor to operate from the Earth's orbit. With 242 spectral bands laid out between 0.4 and 2.5 µm, the sensor enables the extraction of a practically seamless detailed reflectance spectrum for each scene element (pixel) in the image (Goodenough et al., 2002; Galvão et al., 2005).

On the other hand, the spectral response of a pixel is, in fact, the integrated sum of the spectral response of the scene components as registered in the field of view of the sensor (Instantaneous Field of View - IFOV) such as those of the soil, vegetation, water and of the shade projected on the land, as well as atmospheric contributions (Shimabukuro and Smith, 1991).

In order to identify the ratio of the different materials that comprise a pixel, one can use models which decompose the pixel into its pure components or endmembers. The Spectral Mixture Analysis (SMA) (Smith et al., 1985) is a digital image-processing tool that can separate the relative ratio of each material within a pixel by using a set of pure components of the image (Pereira et al., 1998). Nevertheless, the SMA fails in that it considers all pixels to be a mixture of a single initial set of endmembers, thereby possibly allowing a pixel to be modeled by endmembers which might not actually be present therein.

In order to correct such error, the Multiple Endmember Spectral Mixture Analysis model (MESMA) (Roberts et al. 1998a) was proposed as it allows the number and type of endmembers, as well as their prevalence, to vary from pixel to pixel and establishes the best-fit mixture model for each pixel individually.

In this context, the objective of this research was to assess the potential of the data deriving from the Hyperion orbital sensor and from the MESMA model for the discrimination of land cover classes in the Coastal Plains of Rio Grande do Sul.

* Corresponding author

2. MATERIAL AND METHODS

This study used an EO-1 Hyperion stock image, level L1R (USGS), which was acquired on May 2nd, 2004 and depicts the northern area of the Coastal Plains of Rio Grande do Sul (Figure 1).

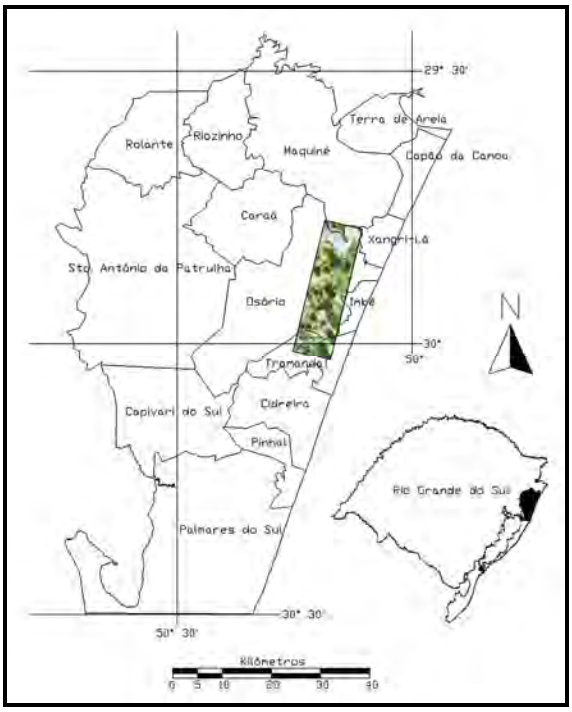


Figure 1. Location of the study area

In order to identify the type of materials present in the scene and characterize land cover classes and subclasses, two two ground truth measurements took place in 2007.

The stages of the preprocessing of Hyperion data involved: correction of anomalous pixels (stripes) by interpolation; exclusion of the bands which had no data of the L1R product and which were positioned around the atmospheric absorption bands (1,400 nm and 1,900 nm); conversion of data into surface reflectance values through the use of a radiative transfer model based on the MODTRAN-4; and image georeferral using a GeoCover image of the study area as the reference.

The Minimum Noise Fraction (MNF), the Pixel Purity Index (PPI) and the n-Dimensional viewer processing sequence was applied over the 143 bands selected from the 454 to 2334 nm interval in order to compose the first spectral library and generate a group of samples for the validation of the results obtained by MESMA.

For the selection of the most representative spectra of the first spectral library, the following metrics were used: EAR (Endmember Average RMSE), MASA (Minimum Average Spectral Angle) and CoB (Count Based Endmember Selection), all included in the VIPER (Visualization and Image Processing for Environmental Research) Tools program (Roberts et al., 2007). The spectra that had the best performance in each metric were selected to generate the second spectral library.

From this second spectral library, the two, three and four endmember MESMA models were applied to the Hyperion data using the "Run SMA/MESMA" module of VIPER Tools. Restriction parameters required by the program regarding the type of shade (photometric or non-photometric), minimum and maximum fractions of the non-shade component, maximum shade fraction, and maximum RMSE allowed are all shown on Table 2.

Shade	Photometric
Minimum Allowable Fraction	-6%
Maximum Allowable Fraction	106%
Maximum Allowable Shade Fraction	80%
Maximum Allowable RMSE	0,025
Residual Threshold	0,025
Number of Contiguous Bands	7

Table 2. Parameters used in the "*Run* SMA/MESMA" module of the VIPER Tools for the application of the MESMA models.

In order to generate the fraction image for each subclass represented in this study as well as a classified image, a MatLab routine was developed, in which the subclass with the greatest ratio was attributed to the pixel.
The accuracy assessment was obtained in relation to the "pure" pixel demixing (Validation Sample) resulting from the PPI process. Thus, the fractions produced by each model for each subclass were compared in relation to the RMSE produced so as to assess the performance of the MESMA mixing models.

3. RESULTS AND DISCUSSION

3.1. Components of the Scene

Lacustrine wetlands are the most present elements in the study area (Figure 3), accounting for approximately 50% of it. Basically, three distinct shades of water were identified and labeled in this study as follows: water with sediments, water without sediments and water with chlorophyll. Current land use is dominated by rice crops and pinus and eucalyptus plantations in small and mid-sized properties. The open field areas are composed of low vegetation, namely grasslands, identified by the bright green shades located among the lacustrine wetlands spread across the image. The dry field areas are characterized by disperse dry vegetation made up of dry pinus branches and rice straw. Finally, a small lot of the urban area of Tramandaí can be seen between the Lagoa das Custódias and Lagoa do Armazém.

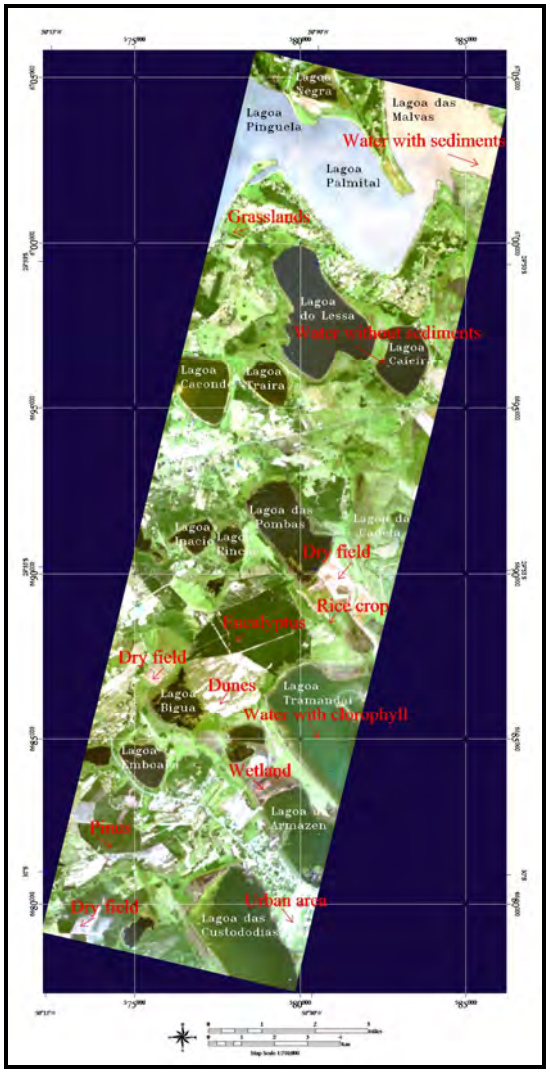


Figure 3. Normal color composition with bands positioned at 638 nm (red), 546 nm (green) and 465 nm (blue). Main components of the scene are highlighted

3.2. First Spectral Library and Validation Sample

In total, 14,101 pixels were selected as endmember candidates by the PPI technique (6.12 % of the imaged area), applied over the 9 first MNF components. Those pixels occurred especially in portions of the scene of 8 subclasses related to three main classes (water, vegetation and soil): water with sediments, water without sediments and water with chlorophyll (water); pinus, eucalyptus and grasslands (vegetation); and dunes and dry fields (soil).

The number of "pure" pixels selected manually from the Hyperion image, with the aid of the n-Dimensional viewer to compose the first spectral library and the validation sample is shown on Table 4.

Classes	Subclasses	Number of pixels (1st Spectral Library)	Number of pixels (Validation Sample)
Water	water with sediments	105	80
	water without sediments	105	80
	water with clorophyll	105	80
Vegetation	pinus	105	80
	eucalyptus	105	80
	grasslands	105	80
Soil	dunes	105	80
	dry fields	20	80
Total of pixels		755	640

Table 4. Classes, subclasses and number of pixels selected to compose the First Spectral Library and the Validation Sample, resulting from the sequential application of MNF, PPI and n-Dimensional viewer techniques

3.3. Second Spectral Library

Using the EAR, MASA and CoB metrics, 23 spectra were selected from the first spectral library. Those endmembers corresponded to the most representative spectra of each subclass, that is, those which presented the lowest RMSE (EAR) and spectral angle and the highest CoB index when modeling the same subclass.

The spectra which presented the best performance in each metric (EAR, MASA and CoB) were selected for each subclass. Therefore, each subclass received three representing spectra (endmembers), except the dune subclass, which, due to the low CoB index produced, had its spectrum selected by the EAR and MASA techniques only.

3.4. MESMA spectral mixture models

The performance result of the 2, 3 and 4 endmember models generated from the second spectral library (Table 5) can be observed in Figure 6. The 4 endmember models obtained the lowest RMSE and the largest fraction when modeling "pure" pixels in the validation sample.

Number of endmembers	1st endmember	2nd endmember	3rd endmember	4th endmember	Number of models
2	water/vegetation/soil	shade			23
3	water/soil	vegetation	shade		126
4	soil	vegetation	water	shade	405

Table 5. Number of MESMA models generated from the 2nd spectral library for performance assessment.

* Corresponding author

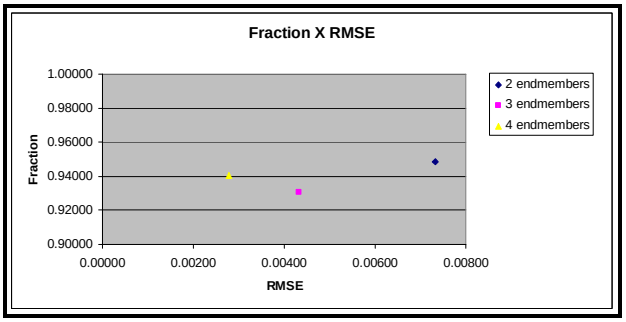


Figure 6. Result obtained by the 2, 3 and 4 endmember MESMA models on the validation sample

Approximately 85% of the Hyperion image was modeled by the 4 endmember MESMA, which corresponds to 195,725 pixels (a total of 230,400 pixels). Most pixels of the water, dry field, pinus, grasslands and eucalyptus subclasses were properly modeled and classified. The dune subclass only had 46.25 % of its sample modeled (Table 7).

Classes	Subclasses	% modeled
Water	Water with Sediments	100,00%
	Water without Sediments	100,00%
	Water with Clorophyll	100,00%
Vegetation	Pinus	100,00%
	Eucalyptus	100,00%
	Grasslands	97,50%
Soil	Dry Fields	100,00%
	Dunes	46,25%

Table 7. Result obtained by the 4 endmember MESMA model on the validation sample

The RMSE image indicated that the most significant mistakes were produced by MESMA when modeling grasslands and dunes due to the high spectral variation of those targets. With the aid of histograms of the fraction images of each subclass, it was found that most of the pixels of the Water Body Class were modeled with fractions superior to 90%. In the Vegetation Class, the average fractions produced varied between 60% (grasslands) and 90% (pinus and eucalyptus). The Soil Class presented the smallest fraction values, with approximately 50% for the dry field subclass. The dune subclass had a practically linear distribution, with intervals between 20% and 104%.

The final classification (Figure 8) obtained from the land vegetation and water fraction images depicts the spatial distribution of the eight subclasses presented in this study. It was observed that some areas that actually corresponded to the cultivation of eucalyptus were erroneously classified as pinus, yet the opposite did not occur. For the water classification, the result seems to be consistent with the normal color image in Figure 3. Nevertheless, in shallow waters (margins), there was a slight confusion between the subclasses Water with Sediments and Water with Chlorophyll, mainly due to the background effect. Similarly to what happened in the validation samples, dune-covered areas were mostly not modeled. The urban area, which was not represented in the mixing models, was partially modeled as dry field (soil).

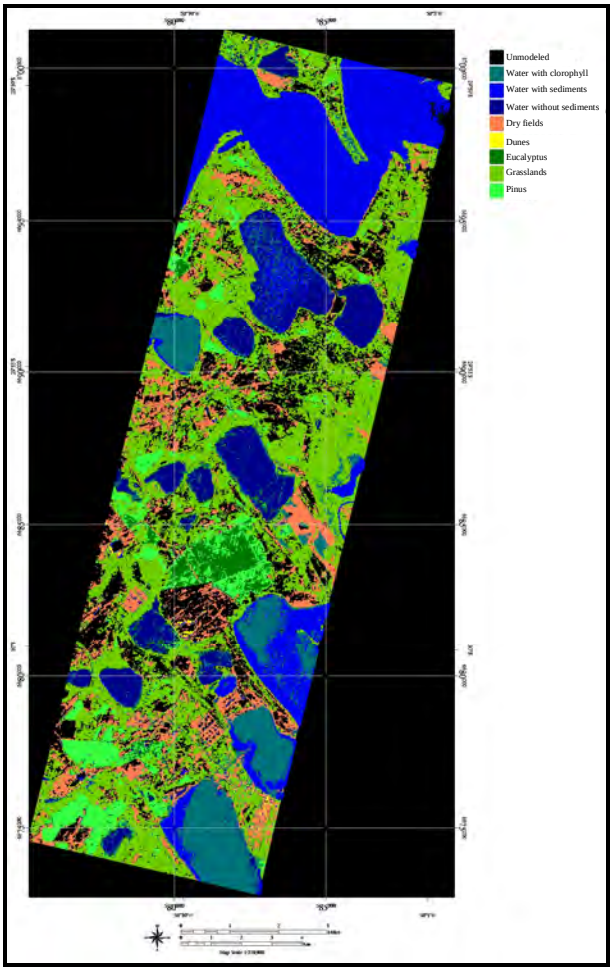


Figure 8. Result obtained by the 4 endmember MESMA model. Image classified from the component with the highest fraction

4. CONCLUSIONS

When applied over a Hyperion image and over the validation sample, the 4 endmember MESMA model (soil = dunes and dry field; green vegetation = pinus, eucalyptus and grasslands; water = with sediments, without sediments and with chlorophyll; shade) properly described the diversity of the scene components, including that of materials within the same class (e.g.: pinus and eucalyptus).

The Fractions produced by the MESMA models when modeling "pure" pixels corresponded to the expected values, mostly reaching fractions above 90% with low RMSE.

The results, as a whole, demonstrate the potential of the application of the MESMA model with EO-1 Hyperion sensor hyperspectral data, even considering the low signal-to-noise relationship of the instrument, especially in the SWIR.

REFERENCES

Galvão, L. S.; Formaggio, A. R.; Tisot, D. A. Discrimination of sugarcane varieties in southeastern Brazil with EO-1 Hyperion data. *Remote Sensing of Environment*, v. 94, p. 523-534, 2005.

Goodenough, D.G.; Bhogal, A.S.; Dyk, A.; Hollinger, A.; Mah, Z.; Niemann, K.O.; Pearlman, J.; Chen, H.; Tan, T.; Love, J.; Mcdonald, S. Monitoring forest with Hyperion and ALI. *IEEE Transactions on Geoscience and Remote Sensing*, v. 2, p. 882-885, 2002.

Pereira, J.L.G.; Batista, G.T.; Thales, M.C.; Roberts, D.A. 1998. Proporção de componentes derivadas de imagens Landsat TM para caracterização da cobertura da terra na amazônia. (CD ROM). *In*: Simpósio Brasileiro de Sensoriamento Remoto, Santos.CD ROM Anais.

Roberts D.A.; Gardner M.; Church R.; Ustin S.; Scheer G.; Green R.O. Mapping chaparral in the Santa Monica Mountains using multiple endmember spectral mixture models. *Remote Sensing of Environment*, v. 65, n. 3, p. 267-279, 1998a.

Roberts, D. A.; Halligan, K.; Dennison, P. VIPER Tools User Manual. Versão 1.2. Santa Barbara: University of Califórnia Santa Barbara, 2007 p. 95. Disponível em: <http://www.vipertools.org> accessed: 18 abr. 2007.

Shimabukuro, Y.E. & Smith, J.A. The least-squares mixing models to generate fraction images derived from remote sensing multispectral data. *IEEE Transactions on Geoscience and Remote Sensing*, v. 29, n.1, p. 16-20, 1991.

Smith, P. S.; Balonek, T. J.; Heckert, P. A.; Elston, R.; Schmidt, G. D. UBVRI field comparison stars for selected active quasars and BL Lacertae objects. *Astronomical Journal*, v. 90, p. 1184-1187, 1985.

URBAN ROAD TRACKING BY FUSION OF SVDD AND REGION ADJACENCY GRAPHS FROM VHR IMAGERY

Z. Liu[a, *], X. Lin [a,b], J. Zhang [a] , P. Pu [a]

[a] Key Laboratory of Mapping from Space of State Bureau of Surveying and Mapping, Chinese Academy of Surveying and Mapping, No.28, Lianhuachixi Road, Haidian District, Beijing 100830, China - (email:zjliu @casm.ac.cn)
[b] School of Resources and Environment Science, Wuhan University, No.129, Luoyu Road, Wuchang District, Wuhan 430079, China

KEY WORDS: Vector, Data, Tracking, Fusion, Imagery

ABSTRACT:

Road surfaces are seriously disturbed by a variety of noises on the very high resolution (VHR) remotely sensed imagery in urban areas, e.g., abrupt geometric deformation and radiometric changes caused by sharp turning, shadows of tall buildings, and appearance of vehicles, which leads to frequent failures for most of current road tracking methods. In this paper, a semi-automatic method is proposed for urban road tracking on VHR imagery. Initially, a human operator inputs three seed points on a selected road, and then necessary information, such as road direction, road width, start point, and a reference template, is automatically derived. The automatic tracking is consequently triggered. During the process, the reference template is moved to generate several target templates. For each target template, a binary template is derived by classifying the target template using support vector data description (SVDD). Subsequently, region adjacency graphs (RAG) is used to eliminate the small disturbing features on the road surfaces in each binary template, which is helpful to search the optimal road centerline points. The above tracking process is repeated until a whole road is completed. Two VHR images were used for the test. The preliminary results show that our method can extract roads more robustly than existing least-squares template matching method in urban areas.

1. INTRODUCTION

The increasing availability of commercial very high resolution (VHR) satellite imaging sensors such as QuickBird, GeoEye-1 and TerraSAR, demands the availability of suitable automatic interpretation tools to extract and identify cartographic features (Lin et al., 2009). Roads are one of the most important cartographic features, and automatic extraction of them is meaningful for various applications such as Geographic Information System (GIS) database updating, transportation analysis and urban planning (Huang and Zhang, 2009). Nevertheless, attempts on developing fully automatic road extraction method for VHR digital imagery have been made for decades (e.g. Hinz and Baumgartner, 2003; Song and Civco, 2004; Jin and Davis, 2005). It still involves several major scientific and technical challenges (Mena, 2003). Therefore, despite a lot of research work on fully automatic approaches for road extraction, the desired high level of automation could not be achieved by now and even in the near future (Baumgartner et al., 2002). One more practical solution to this problem is to adopt a semi-automatic approach that retains the "human in the loop" where the computer vision algorithms are used to assist human extracting the roads (Zhou et al., 2006). Currently, dozens of semi-automatic methods are proposed for road extraction from VHR imagery, and many of them reach various levels of success. In general, these semi-automatic approaches may be grouped into two categories: path optimizers and road trackers or path finders (Amo et al., 2006).

A path optimizer is applied to determine an optimal trajectory between manually selected seed points, and it is often realized by improving the dynamic programming and snakes or active contour model (Gruen and Li, 1997) for VHR images. In these models, geometric and radiometric characteristics of roads are integrated by a cost function or an 'energy' function, and then the road extraction is equivalent to seeking the global energy minimum. Amo et al. (2006) improved the active contour model by the region competition algorithm to extract the ribbon roads on aerial images. Dal Poz and do Vale (2003) made a modification of merit function of the original dynamic programming approach, which is carried out by a constraint function embedding road edge properties. However, it is hard to define the reasonable 'energy' function for each road on each VHR image.

Compared to the path optimizers, path finders are more popular. A path finder is an iterative line growing process: it often starts with some seed points, then the local information is used to add new segments into the road network based on the pixel intensities of the image, and typically a human operator is needed to help the path finder go through the various types of image noises such as cars and shadows. For example, McKeown and Denlinger (1988) described one of the most general road finder based on the cooperation between the intensity profile correlation of road cross sections and road edges following. Vosselman and Knecht (1995) imposed the profile matching by using least squares template matching and Kalman filter. Baumgartner et al. (2002) also presented a human-computer interactive prototype system by the above method. Similarly, Zhou et al. (2006) used two profiles, one perpendicular to road direction and the other parallel to road direction, to enhance robustness of the tracker and applied extended Kalman filter and particle filter to solve profile matching issues for road tracking. Slightly different from the above methods; Kim et al. (2004) employed a rectangular reference template of road surfaces to track roads by least squares template matching, and road path is modelled as similarity transform; Hu et al. (2004) presented a road finder using a piecewise parabolic model and least-squares template matching; Zhou et al. (2007) utilized one on-line learning

* Corresponding author

method based on the principle of one-class support vector machines (SVMs) to find the optimal matched template in road tracking; and Lin et al. (2009) described a road finder by both tracking the lane markings and road surfaces based on least-squares template matching. However, most of the above road finders fail when they encounter the road intersections. Another road finder presented by Hu et al. (2007) can well extract the intersections besides the general roads, and it employed a spoke wheel operator to obtain the road footprints. Despite most of the above road finders perform well on some kinds of roads or intersections such as highways or rural roads where the road surfaces are relatively homogenous on VHR images, they often failed to extract the roads where the surfaces suffer from abrupt geometric deformation and radiometric changes caused by sharp turning, shadows of tall buildings, and appearance of vehicles etc.

How to decrease the negative effects of various types of image noises is a key step to increase the robustness of a path finder. Actually, most of existing path finders, such as McKeown and Denlinger's one (1988), make use of least-squares template matching in searching an optimal road centreline point, but this type of method, using the squared sum of the grey value differences as a measure, is easily impacted by the image noises mentioned above. In this case, new features of roads should be selected and utilized. For example, Zhao et al. (2002) utilized the template matching on a classified image in a semi-automatic road tracking system, and Lin et al. (2009) proposed a novel road signature measure called "parallelepiped angular texture signature" to semi-automatically track roads based on the unique characteristic of roads on a classified sub-windows. It is testified that the supervised classification can indeed provide a novel feature for road tracking. However, most of the existing conventional supervised classification analyses may depict multiple classes including buildings, water, trees etc. besides the roads and they assume implicitly that the set of training sample for each class is large enough (Foody et al., 2006). However for road tracking application, our interest is only focused on just one specific class, road, and the training set size is not large enough in road mapping. Recently, statistical learning theory and one-class SVMs have been used in road extraction from VHR images, e.g., Zhou et al.'s method (2007) mentioned above. The support vector data description (SVDD) is a one-class classifier based on the principles of the SVM, and it provides a very simple to use supervised classification analysis that requires only the training data for the class of interest (Sanchez-Hernandez et al., 2007). Moreover, the accuracy of SVDD classification was considerably higher than that derived from a conventional multi-class parametric classification (e.g., Maximum Likelihood) and popular alternatives (e.g. feedforward neural networks) (Sanchez-Hernandez et al., 2007).

In this paper, SVDD is employed to track road on VHR images for the first time. Particularly, once training samples are provided from the reference template derived from road initialization, SVDD is trained and used to identify road pixels in sub-windows generated in moving of reference template, and, subsequently region adjacency graphs (RAG) is employed to eliminated the image noises contained by road surfaces in the classified sub-windows, and template matching is utilized to determine the optimal road direction for road tracking.

2. RELATED COMPUTER ALGORITHMS

One-Class Classification by SVDD

Road is the only specific class of interest in road mapping from VHR imagery, and a range of approaches exist to classify a specify class of interest, including reconstruction methods (e.g. Pizzi et al., 2001), density methods (e.g. Fumera et al., 2000), and boundary methods (e.g. Zhou et al., 2007). However, reconstruction methods and density methods require extensive knowledge and large amount of information about the data set of interest. Fortunately, boundary methods are more feasible in that they do not require the extensive knowledge of the data set, as they concentrate on the boundary that fits around the class of interest (Tax, 2001). This benefit makes the boundary methods very attractive to use in remote sensing applications (Sanchez-Hernandez et al., 2007). Boundary methods are largely based the statistical learning method (Vapnik, 1995) and the principles of SVMs (Song and Civco, 2004), and the recently developed SVDD is comparable to SVMs. The basic idea of SVMs binary classifier that seeks to fit an optimal separating hyperplane or decision boundary between the classes; however, the SVDD searches for a closed boundary around the training data, namely a hypersphere, instead of looking for a hyperplane (Tax and Duin, 2004).

The hypersphere may be defined by $F(R,O) = R^2$, where O is the centre and R is the radius. Therefore, the problem SVDD attempts to solve is to find the hypersphere with the constraints that all the training data x_i are within R^2. Figure 1 shows the geometrical interpretation in a two dimensional case. The problem can be formulated as follows:

$$\text{Minimize } (F(R,O) = R^2)$$
$$\text{subject to } \| x_i - O \| \le R^2, \qquad \forall_i. \qquad (1)$$

In order to allow the possibility of outliers in the training set, the distance from x_i to the center O should not be strictly smaller than R^2 but larger distance should be penalized. In this sense, slack variables $\varepsilon_i \ge 0$ must be introduced into the error function and, correspondingly, the above optimization problem changes into:

$$\text{Minimize } (F(R,O,\varepsilon) = R^2 + C\sum_i \varepsilon_i)$$
$$\text{subject to } \| x_i - O \| \le R^2 + \varepsilon_i, \quad \varepsilon_i \ge 0, \quad \forall_i \qquad (2)$$

where C is a known and given coefficient that makes a trade-off between the volume of the description and the misclassification errors.

Using the technique of Lagrange multipliers, this optimization problem can be formulated into the following quadratic programming problem:

$$\underset{\alpha,\gamma}{Max}\ \underset{R,O,\varepsilon}{Min}\ L(R,O,\alpha,\gamma,\varepsilon) = R^2 + C\sum_i \varepsilon_i - \sum_i \alpha_i \times \{R^2 + \varepsilon_i - (\| x_i \|^2 - 2x_i \cdot O + \| O \|^2)\} - \sum_i \gamma_i \varepsilon_i$$

subject to $\alpha_i \geq 0$, $\gamma_i \geq 0$, $\sum \alpha_i = 1$, $O = \sum_i \alpha_i x_i$, C -

$$\alpha_i - \gamma_i = 0, \quad \forall_i . \tag{3}$$

where α_i, γ_i are Lagrange multipliers respectively, and $\underset{\alpha,\gamma}{Max}\,\underset{R,O,\varepsilon}{Min}$ means $L(R,O,\alpha,\gamma,\varepsilon)$ should be minimized with respect to R, O, ε and maximized with respect to α, γ. Substituting the last three constraints into the target function will give the following simplified formula:

$$\text{Maximize } L = \sum_i \alpha_i (x_i \cdot x_i) - \sum_{i,j} \alpha_i \alpha_j (x_i \cdot x_j)$$

$$\text{subject to } 0 \leq \alpha_i \leq C, \sum \alpha_i = 1, \quad \forall_i \tag{4}$$

Note that the magnitude of the Lagrangian multiplier α_i varies with the position of the case relative to the hpersphere. Figure 1 shows a case within the hypersphere where α_i=0; a case on the hypersphere boundary where 0< α_i < C; and a case outside the hypersphere where $\alpha_i = C$. Moreover, only the samples with 0< α_i < C are the support vectors of the description (Sanchez-Hernandez et al., 2007), which are essential for the calculation of the optimal hypersphere with centre O and radius R.

The solution of SVDD is given by:

$$O = \sum_i \alpha_i x_i . \tag{5}$$

While the decision function for the SVDD classification is given by:

$$f(x) = \| x - O \|^2 \leq R^2 \tag{6}$$

For the non-linear case, as with SVMs, noticing the training data appeared in the optimization problem in the form of dot products, a mapping ϕ of the data using the kernel functions (Commonly-used kernel functions refer to Song and Civco, 2004) may be firstly denoted as:

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) . \tag{7}$$

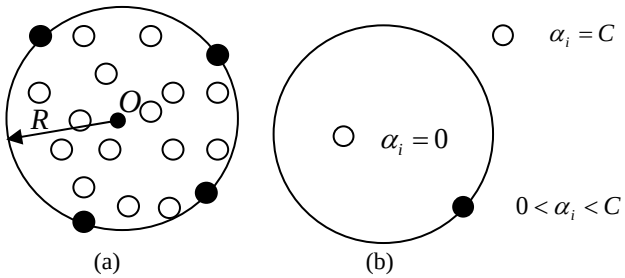


Figure 1. Support vector data description. (a) Hypersphere containing the target data. The shaded objects on the edge of the sphere are the support vectors. (b) Magnitude of the two Lagrange multipliers for cases inside, on the boundary, and outside the hypersphere.

In this paper, a human operator will provide the training data for the SVDD classifier by a reference template, as being introduced in the next section, and the classification will be performed on a patch of the image. Moreover, in the SVDD classification, the polynomial-degree function kernel of free parameter value two and C = 0.01 was selected as done by (Sanchez-Hernandez et al., 2007).

Noise Removal by RAG

The above SVDD classification procedure produces a patch of binary image, which labels pixels belonging to road class as 1 while the other non-road pixels as 0. Some parts of road surfaces may be misclassified into non-road class due to the various types of image noises such as occlusions of vehicles, shadows of trees and buildings, as shown in Figure 3. If we suppose that any image primitives, belonging to non-road class, contained in road class polygons are road class primitives, these primitives should be reclassified into road class. Herein, RAG, as shown in Figure 3, is employed to do the topological analysis and reclassify the noises on road surfaces into road class, which will significantly decrease the side-effects of the image noises. Figure 3 illustrates that the road surfaces are dilated and the noises are eroded after the RAG analysis, which decrease the negative effects of noises.

3. SYSTEM FRAMEWORK

In our semi-automatic system, a human operator is required in the road extraction process where computer algorithms are utilized to assist the operator performing measurement tasks. From the user's point of view, the procedure is as follows: the operator first inputs three seed points that detect a short segment of a road which serves as initialization for computer algorithms, and then the proposed algorithms are launched and automatically track the road axis as long as possible. Whenever the internal evaluation of the algorithms indicates that the tracker might have lost the road centreline, the system needs intervention of the user. Then the operator has to confirm the path finder (tracker) or he/she must edit the extracted road and put the tracker back the road again. Concretely, the system is based on the following road model and the main procedures.

Suppose a road model is represented as a queue of road centreline points that is denote as:

$$\{(x_0, y_0, \theta_0), (x_1, y_1, \theta_1), \cdots, (x_i, y_i, \theta_i), \cdots, (x_n, y_n, \theta_n)\}$$

where (x_i, y_i) are the planar coordinates of the *i*th road centreline point while θ_i is the corresponding direction of the above road point, and the relationship between the *i*th point and the (*i*-1)th point can be expressed as:

$$\begin{bmatrix} x_i \\ y_i \\ \theta_i \end{bmatrix} = \begin{bmatrix} x_{i-1} + L_{step} * \cos\theta_{i-1} \\ y_{i-1} + L_{step} * \sin\theta_{i-1} \\ \theta_{i-1} \end{bmatrix} (i \geq 1) \tag{8}$$

where L_{step} is the suitable step size of the increment (i.e., the distance between two consecutive points on the road axis), and L_{step} is set to road width in this paper. As a result, if (x_0, y_0, θ_0) is known, then the automatic road tracking is equivalent to searching the optimal direction for each road point. Particularly, the semi-automatic tracking is divided into the following steps.

Step 1: Initialization by three seed points

Similar to Vosselman and Knecht's method (1995), the initialization is also accomplished by manually selected seed points. However, we take another strategy in which a three consecutive mouse clicks strategy is adopted to obtain the starting point, direction, width of the road, and the step size as well. This three seeds method is feasible for most of the roads on VHR images, and it is accomplished as follows (see Figure 2): the human operator enters a road segment with two

consecutive mouse clicks on A' and B with the axis joining the points defining one road sideline A'B, which indicates road 9direction arctangent(A'B), then the following third click on C, on the other roadside, defines the width w of the road. w is equal to the distance between the point C and the line A'B. As a result, the above three points can also derive a rectangle A'B'B"A" with width w and length l. Particularly, the direction of A'B' is equal to arctangent(A'B) while C is located on the side B"A", and l = 2* w. Then a start point A, denoted as (x_0, y_0), of the road is derived from the middle point of A' A". The above information forms the first road centreline point (x_0, y_0, θ_0) in the road model mentioned above where θ_0 = arctangent(A'B), and obtain the next road centreline point (x_1, y_1, θ_1) by Equation 8 where $\theta_1 = \theta_0$, add the above two points into the road queue sequentially. Simultaneously, the pixels in the template A'B'B"A" also serves as training samples for the SVDD classifier, and then a predict model is derived by training the SVDD classifier.

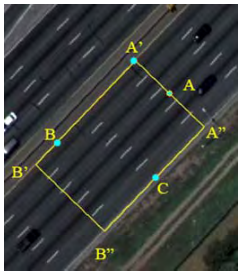


Figure 2. Road initialization by three seed points

Step 2: Acquire the next road axis candidate point

Take the two latest road centreline points out from the current road queue, denoted as $(x_{i-1}, y_{i-1}, \theta_{i-1})$ and (x_i, y_i, θ_i), respectively. Revolve around the pixel $p\ (x_i, y_i)$, and form a square $S(\ p(x_{TopLeft}, y_{TopLeft}), p(x_{BottomRight}, y_{BottomRight})\ )$, where $p(x_{TopLeft}, y_{TopLeft})$ and $p(x_{BottomRight},\ y_{BottomRight})$ are the top left corner and bottom right corner of the square respectively. Concretely, the coordinates of two corners are calculated by the following formula:

$$\begin{aligned} x_{TopLeft} &= x_i + \sqrt{2} * l * \cos(\theta_{i-1} - \frac{3}{4}\pi) \\ y_{TopLeft} &= y_i + \sqrt{2} * l * \sin(\theta_{i-1} - \frac{3}{4}\pi) \\ x_{BottomRight} &= x_i + \sqrt{2} * l * \cos(\theta_{i-1} + \frac{3}{4}\pi) \\ y_{BottomRight} &= y_i + \sqrt{2} * l * \sin(\theta_{i-1} + \frac{3}{4}\pi) \end{aligned} \quad (9)$$

where l is the length of the reference rectangle introduced in the first step.
As mentioned above, the reference A'B'B"A" derives a predict model for the SVDD classifier, and then perform SVDD classification on the above obtained squared subset image. Subsequently, set the pixels of road subclass as 1, meanwhile set the pixels of any other subclass as 0, and then perform the RAG analysis on the binary image to reclassify the image noises on the road surfaces into road class, which will decrease the negative effects of various types of noises.

Following, at each road centreline point $p\ (x_i, y_i)$, a rectangular template with width w and height l is revolved on the classified image, and $T(\alpha, w, h, p)$ is defined as the mean for the rectangular set of pixels of around pixel p whose principal axis lies at an angle of α from the road direction θ_{i-1}. This measure is computed for a set of angles $\alpha_0, \ldots \alpha_n$ at pixel $p\ (x_i, y_i)$. Angles $\alpha_0, \ldots \alpha_n$ are with same interval δ. At the point p, the mean of the template at each rotating angle forms a set of values $\{T(\alpha_0, w, h, p), T(\alpha_1, w, h, p), \ldots, T(\alpha_n, w, h, p)\}$, named as classified angular texture signature (CATS). Figure 3(a) shows a CATS with δ =5°. The direction of the significant maximum which has a minimal inclination with road direction θ_{i-1} is taken as the real direction of current road axis point, and replace θ_{i-1} with the optimal value.

Step 3: Validate the above optimal point
Once the above obtained point is added into the road model, check whether any stopping criterion is fulfilled as follows:

- the change of the directions of two adjacent road segments is larger than predefined threshold T;
- the minimal mean value of the optimal template surpass T_1;
- compactness of CATS polygon is larger than T_2;
- approaching an extracted road or border of the image.

To find the relationship between the shape of the CATS polygon and corresponding pixel types, we plot the CATS values around the pixel under consideration with corresponding direction and link the last point to the first point. The resulting polygon is called the CATS polygon, and Figure 3(e) shows the calculated CATS for pixel p with the CATS polygons. If the road has a good contrast with its surrounding objects, the polygon usually looks like an ellipse or ∞-shape, or a circle in other cases. The compactness of CATS can be defined as the compactness of the CATS polygon using Equation (10):

$$\mathrm{CATS}_{\mathrm{compactness}} = \frac{4\pi \cdot A}{P^2} \quad (10)$$

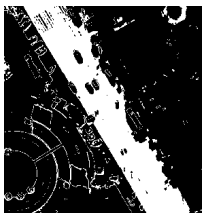
where *A* and *P* are the area and perimeter of the CATS polygon, respectively. It is employed to check whether the shape of the CATS polygon looks like a circle. A circle-like CATS polygon usually indicates that the tracker is no longer fit for tracking the road ahead. Note that our program will calculate the compactness of CATS at regular intervals to verify whether the CATS is still suitable for tracing a road.
If any of these conditions is encountered, exit the tracking procedure and go to Step 4). Otherwise, obtain the next road centreline point $(x_{i-1}, y_{i-1}, \theta_{i-1})$ by equation (8) and add this point into the road queue, and go to Step 2 again.
Step 4: Stop the automatic following

If no rule can be made to continue the tracking procedure, the system will stop tracking, report the reason, and offer an appropriate choice of user interaction. The user can then modify the traced path with the aid of common GIS-functionalities, manually digitize complex roads, update the reference template (occurrence of change of the number of lanes, or significant change of spectral characteristics due to

different ages, construction materials, illumination angles, etc.), or restart the tracking process from the next specified location.

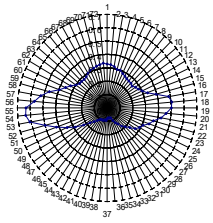
(a) A rotating rectangular template and its resulted sub-windows

(b) Resulted sub-window

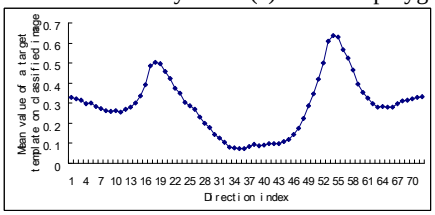


(c) Classified image by SVDD

(d) Image noises removal by RAG



(e) A CATS polygon



(f) Values of the CATS in image (a)

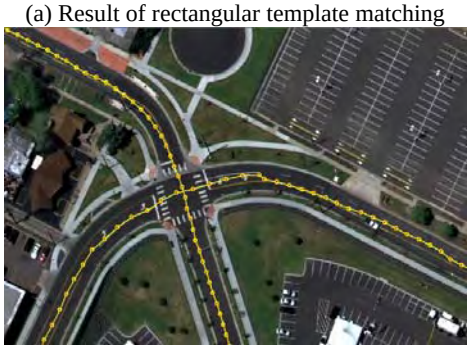
Figure 3. Road direction determination by SVDD and RAG

4. EXPERIMENTS AND PERFORMANCE EVALUATION

A prototype system, based on our proposed method and rectangular template matching (Kim et al. 2004) is developed in VC++6.0 IDE under Win-XP OS. Note that the implementation SVDD is based on Tax's code in matlab environment (Tax, 2001) and the standard one-class SVM in LibSVM (Chang and Lin, 2001).

Two airborne images in urban areas were tested to verify the capabilities of each road tracker. The roads on the above two images are disturbed by various types of image noises such as zebras, occlusions of vehicles, material change, and the extracted results are shown in Figure 4 and Figure 5, respectively. For the first image, the existing rectangular template matching method is feasible, but it failed at the sharp turning and the intersection, and it also failed to track the ring road around the stadium in the second image due to large change of radiometric characteristic of the road. Fortunately, the proposed method succeeds to extract the accurate centrelines of roads in the above two images. The above two tests suggest that our proposed method is more robust to various types of image noises such as sharp turnings, road intersections, zebras, vehicles and material change etc.

(a) Result of rectangular template matching



(b) Result of our method

Figure 4. Extracted roads at an intersection

5. CONCLUSIONS

This paper presents a semi-automatic system for road tracking from VHR remotely sensed imagery. Once a human operator input three seed points that derive a reference template, the system adopts a new combination strategy to automatically track the road networks. Particularly, in the automatic tracking process, SVDD classifier is employed to produce a patch of classified binary image based on the reference template, RAG is utilized to erode the various types of image noises and enhance the road feature space on the binary image, and template matching using mean of the values of the pixels in a target template is used to search the optimal road direction and next road centreline point. The above procedure is repeated until a whole road is tracked. At the same time, a human operator is retained in the tracking process to supervise the extracted results, to response to the program's prompts. Experiments are performed to extract roads from aerial/satellite imagery. The results show that our proposed road trackers can more robustly extract most of the main roads than other typical road trackers, which have significant practical applications. Future work will also include the optimization of the algorithms to speed up the calculations.

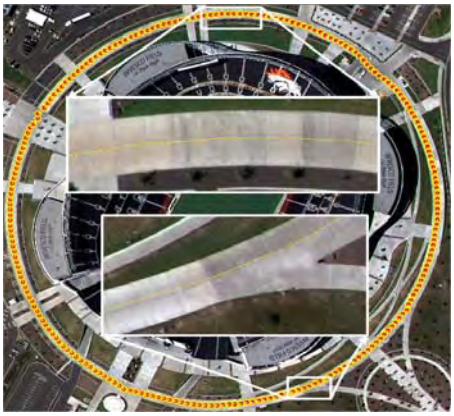


Figure 5. The extracted ring road of a stadium by our proposed method (Note that the rectangular template matching method failed on this image)

REFERENCES

Amo, M., Martínez, F., and Torre, M., 2006. Road extraction from aerial images using a region competition algorithm. *IEEE Transactions on Image Processing*, 15(5), pp. 1192-1201.

Baumgartner, A., Hinz, S., and Wiedemann, C., 2002. Efficient methods and interfaces for road tracking. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 34, Part 3B, pp. 309-312.

Chang, C.C., and Lin, C.J., 2001. LIBSVM: a library for support vector machines. http://www.csie.ntu.edu.tw/~cjlin/libsvm (accessed 20 July 2007).

Dal Poz, A.P., and do Vale, G.M., 2003. Dynamic programming approach for semi-automated road extraction form medium-and high-resolution images. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXIV, Part 3/W8, pp. 87-91.

Foody, G.M., Mathur, A., Sanchez-Hermandez, C., and Boyd, D.S., 2006. Training set size requirements for the classification of a specific class. *Remote Sensing of Environment*, 104, pp. 1-14.

Fumera, G., Roli, F., and Giacinto, G., 2000. Reject option with multiple thresholds. *Pattern Recognition*, 33(12), pp. 2099–2101.

Gruen, A., and Li, H., 1997. Semi-automatic linear feature extraction by dynamic programming and LSB-Snakes. *Photogrammetic Engineering and Remote Sensing*, 63(8), pp. 985-995.

Hinz S., and Baumgartner, A., 2003. Automatic extraction of urban road network from multi-view aerial imagery. *ISPRS Journal of Photogrammetry & Remote Sensing*, 58, pp. 83-98.

Hu, X., Zhang, Z., and Tao, C.V., 2004, A robust method for semi-automatic extraction of road centrelines using a piecewise parabolic model and least squares template matching. *Photogrammetric Engineering and Remote Sensing*, 70(12), pp. 1393-1398.

Hu, J., Razdan, A., Femiani, J.C., Cui, M., and Wonka, P., 2007. Road network extraction and intersection detection from aerial images by tracking road footprints. *IEEE Transaction on Geoscience and Remote Sensing*, 45(12), pp. 4144-4157.

Huang, X., and Zhang, L.P., 2009. Road centreline extraction from high-resolution imagery based on multi-scale structural features and support vector machines, *International Journal of Remote Sensing*, 30(8), pp. 1977-1987.

Jin, X., and Davis, C.H., 2005. An integrated system for automatic road mapping from high-resolution multi-spectral satellite imagery by information fusion. *Information Fusion*, 6, pp. 257-273.

Kim, T., Park, S.R., Kim, M.G., Jeong, S., and Kim, K.O., 2004. Tracking road centerlines from high resolution remote sensing images by least squares correlation matching. *Photogrammetric Engineering and Remote Sensing*, 70(12), pp. 1417-1422.

Lin, X., Liu, Z., Zhang, J., and Shen, J., 2009. Combining multiple algorithms for road network tracking from multiple source remotely sensed imagery: a practical system and performance evaluation. *Sensors*, 9(2), pp. 1237-1258.

Mena J.B., 2003. State of the art on automatic road extraction for GIS update: a novel classification. *Pattern Recognition Letters*, 24(16), pp. 3037- 3058.

Pizzi, N. J., Vivanco, R.A., and Somorjai, R.L., 2001. EvIdent: A functional magnetic resonance image analysis system. *Artificial Intelligence in Medicine*, 21(1–3), pp. 263–269.

Sanchez-Hernandez, C., Boyd, D.S., and Foody, G.M., 2007. One-class classification for mapping a specific land-cover class: SVDD classification of Fenland. *IEEE Transactions on Geoscience and Remote Sensing*, 45(4), pp. 1061-1073.

Song, M., and Civco, D., 2004. Road extraction using SVM and image segmentation. *Photogrammertric Engineering and Remote Sensing*, 70(12), pp. 1365-1371.

Tax, D.M.J., 2001. One-class classification. Ph.D. dissertation, Delft University Technology, Delft, The Netherlands.
Tax, D.M.J., and Duin, R.P.W., 2004. Support vector data description. *Machine Learning*, 54(1), pp. 45–66.

Vapnik, V., 1995. The nature of statistical learning theory. New York: Springer-Verlag.

Vosselman, G., and Knecht, D.J., 1995. Road tracing by profile matching and Kalman filtering. In: *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, A. Gruen, O. Kuebler and P. Agouris (Eds)(Basel, Switzerland: Birkhauser Verlag), pp. 265–274.

Zhao, H., Kumagai, J., Nakagawa M., and Shibasaki, R., 2002. Semi-automatic road extraction from high resolution satellite images. In: *Proceedings of ISPRS Photogrammetry and Computer Vision*, Graz, Ustrailia, pp. A-406.

Zhou, J., Bischof, W.F., and Caelli, T., 2006. Road tracking in aerial images based on human-computer interaction and Bayesian filtering. *ISPRS Journal of Photogrammetry & Remote Sensing*, 61(2), pp. 108-124.

Zhou, J., Li, C., and Bischof, W.F., 2007. Online learning with novelty detection in human-guided road tracking. *IEEE Transactions on Geoscience and Remote Sensing*, 45(12), pp. 3967-3977.

ACKNOWLEDGEMENTS

This research was funded by the National Key Basic Research and Develop Program under Grant 2006CB701303 and the Project for Young Scientist Fund sponsored by the Natural Science Foundations of China under Grant 40401037.

VALIDATION OF THE RADIOMETRIC PROCESSING CHAIN OF THE LEICA ADS40 AIRBORNE PHOTOGRAMMETRIC SENSOR

L. Markelin [a, *], E. Honkavaara [a], U. Beisl [b], I. Korpela [c]

[a] Finnish Geodetic Institute, Masala, Finland - (lauri.markelin, eija.honkavaara)@fgi.fi
[b] Leica Geosystems, Heerbrugg, Switzerland - ulrich.beisl@leica-geosystems.com
[c] Faculty of Agriculture and Forestry, P.O. Box 27, 00014,University of Helsinki - ilkka.korpela@helsinki.fi

KEY WORDS: Camera, Image, Radiometric, Calibration, Correction, Quality, Accuracy, Multispectral

ABSTRACT:

Large-format photogrammetric digital airborne imaging sensors have been commercially available for several years. Their excellent radiometric properties compared to film-based imaging have been reported in several studies. Despite these radiometric advantages, up to now, airborne digital images have been exploited in a rather conventional manner in photogrammetric applications, even though methodologies from quantitative remote sensing, e.g. radiometric image correction and classification, could be utilized to raise the performance of photogrammetric applications to a new level. The recent state-of-the-art review revealed that the fundamental problem in the quantitative utilization of image radiometry in photogrammetric applications is the radiometric correction. Among digital photogrammetric large-format mapping sensors, the Airborne Digital Sensor (ADS) of Leica Geosystems is the only commercially available system at the moment having an integrated, physically based, radiometric correction chain. The processing does not require any in situ control information, as the radiometric corrections are based on *a priori* calibration information and image data. To validate performance of the processing chain, a comprehensive flight campaign was carried out with an ADS40 SH52 sensor in Finland in August 2008. We present the first results of the validation of the Leica ADS40 radiometric processing chain. The results indicated great performance potential. With the challenging data set, the differences of ADS40 and independent ground reference reflectance measurements were even less than 5% for uniform targets; atmospheric state, multispectral channel and flying height were detected as the major factors influencing the accuracy. Leica Geosystem's ADS40 can be considered as an efficient and accurate, 3D, multi-angular, multispectral imaging radiometer, which opens new interesting prospects for 3D remote sensing and characterization of the Earth surface. Results also indicated the importance of the test field validation process, gave improvement ideas for the sensor post-processing software and provided information for the development of validation methods.

1. INTRODUCTION

Digital imaging is replacing film imaging in photogrammetric data capture. In addition to the rigorous 3D-geometric performance, digital photogrammetric sensors offer excellent radiometric potential (Markelin et al., 2008).

A recent assessment of the state-of-the-art of radiometric processing in photogrammetric production lines of several European National Mapping Agencies showed that the radiometry is not processed quantitatively in operational processes (Honkavaara et al., 2009). Instead, radiometric processing is a complicated and laborious task, and the results are not typically satisfactory. The expected benefits of more accurate radiometric processing are the more automatic image processing, higher quality true-color and reflectance imagery, and better automation level of applications. Expected possibilities of the accurate, photogrammetric reflectance data are, for instance, reliable production of vegetation indices, utilization of spectral libraries, time series analyses, and enhanced change detection. In Finland an important prospect is the enhancement of the tree species classification, which is currently the bottleneck in automation of forest interpretation.

The requirements of quantitative radiometry have been taken into account in all aspects of the large-format photogrammetric mapping sensor of the Leica Geosystems, the Airborne Digital Sensor (ADS40, ADS80). The basis is an accurate, stable sensor, which performs radiance measurements in the blue (B), green (G), red (R) and near infrared (NIR) ranges of the electromagnetic spectrum (Fricker, 2007). An important feature is the accurate radiometric laboratory calibration which is applied rigorously throughout the data processing (Beisl, 2006). The radiometry chain is completed by physically based radiometric correction methods, which are implemented in the ADS post-processing software, the Leica XPro (Beisl et al., 2008).

Leica Geosystems was the first photogrammetric sensor manufacturer to integrate quantitative processing of radiometry in the sensor post-processing line (Beisl, 2006; Beisl et al., 2008). Recently, also Intergraph has started improving the radiometric processing of the DMC (Ryan and Pagnutti, 2009). The practical experiences of XPro have indicated great improvements of processing efficiency, better radiometric homogeneity of the output image mosaics and improved performance e.g. in forestry applications (Beisl et al., 2008). However, thus far there exists no independent, quantitative assessment of the performance of the methodology.

The objective of this study was to validate the Leica ADS40/XPro at-sensor radiance and atmospherically corrected ground reflectance products. We also considered radiometric validation aspects of airborne sensors. The study was performed using ADS40 data from a comprehensive empirical flight campaign carried out in Finland in August 2008. The image data set is one of the image materials offered in the context of the European Spatial Data Research (EuroSDR) research

* Corresponding author.

project "Radiometric aspects of digital photogrammetric airborne images" (Honkavaara et al., 2009).

2. MATERIALS AND METHODS

2.1 The Leica ADS40 sensor and the XPro radiometric processing chain

The ADS40 is a large-format photogrammetric sensor based on the pushbroom principle. In this study the sensor head 52 (SH52) is used which has a total of 12 CCD-lines installed in different positions on the focal plane to provide different along track viewing angles: nadir panchromatic (PAN; 2˚), nadir R, G, B and NIR (0˚), backward PAN (14˚), backward R, G, B and NIR (16˚) and forward PAN (27˚) (Fricker, 2007).

The manufacturer performs comprehensive absolute radiometric calibration for the ADS, including corrections for dark signal non-uniformity (DSNU), photo response non-uniformity (PRNU), absolute radiometric response, and spectral response (Beisl, 2006). Since the sensor response w.r.t. incident radiance data and integration time is linear, the radiometric model for retrieving the at-sensor radiance from raw digital numbers (DN) for a specific multispectral (MS) channel is given by:

$$CDN = DN * 50 * c_1 / IT \quad (1)$$

where CDN is calibrated DN (16 bit integer), DN is recorded raw DN (16 bit integer), c_1 is radiometric gain (camera and channel specific), IT is integration time (s) and 50 is scale factor.

CDN data are stored as epipolar rectified 16 bit integer images. The band-averaged spectral radiance L [W/(m^2 sr um)] of the band is calculated by dividing the CDN with the scale factor 50.

Leica XPro is used for the entire post-processing workflow of the ADS-imagery from data download to the generation of stereo models and orthoimages. In radiometric terms, main features are the options to produce radiometrically corrected ground radiance and ground reflectance images. The default product of XPro is calibrated DN (equation 1), which relates the pixel data to at-sensor radiances. There are two options to produce ground radiance data: the Dark Pixel Subtraction (Chavez, 1975) and the Modified Chavez (Chavez, 1988) methods. Ground radiances are still dependent on the illumination level and vary from flight line to flight line. To overcome this, there is an option for atmospheric correction and reflectance calibration based on the radiative transfer equation by Fraser et al. (1992) and a parameterization of the atmospheric parameters based on the method of Song et al. (2003). This atmospheric correction results in images where the DNs are calibrated to ground reflectances. In ground reflectances, the reflected radiance is divided by the incoming solar irradiance which results in a surface property. All three correction methods are based on an automatic dark object method to tune the corrections to the actual atmospheric conditions. Additionally, BRDF (Bidirectional reflectance distribution function) correction based on a modified Walthall model is implemented in XPro. The details of the correction methods are in Beisl et al. (2008). All corrections rely entirely on *a priori* calibration information and atmospheric information derived from dark pixels in the image data.

2.2 Imagery

A flight campaign was carried out at the Hyytiälä forestry research station in Finland (62°N, 24°E) on 23 August, 2008 using a Leica ADS40 SH52 digital photogrammetric camera to validate the sensor performance and to evaluate data performance in forestry applications. A total of 15 flight lines were collected from four flying heights (1, 2, 3 and 4 km, resulting ground sampling distances (GSDs) 10, 20, 30 and 40 cm, respectively) of which 4 were used here. The MS channels (R, G, B, NIR, both nadir and backward directions) were recorded in raw (uncompressed) mode. The weather conditions were mostly clear, but some small clouds were over the area. The detailed information of the images used and atmospheric conditions are in Table 1.

Flying height	1km	2km	3km	4km
GSD [cm]	10	20	30	40
Flying direction	349°	349°	349°	169°
Integration time [ms]	1.94	2.77	4.16	5.54
Off-nadir viewing angle	15°	3.5°	3.2°	10.7°
Start time (UTC+3)	10:25	10:45	11:18	11:43
End time (UTC+3)	10:28	10:48	11:20	11:47
Sun elevation angle	30.0°	31.8°	34.5°	36.2°
Sun azimuth angle	126.6°	131.9°	141°	148.3°
Visibilty [km]	49.6	42.8	50.0	50.0
Temperature [°C]	14.6	14.9	15.0	15.5
AOT 500nm	0.16	0.15	0.17	0.16
CO2 [ppm]	373	372	374	372
O3 [g/cm2]	6.54E-04	6.54E-04	6.54E-04	6.54E-04
H20 [g/cm2]	1.41	1.40	1.45	1.39

Table 1. Information of image lines used and weather conditions.

2.2.1 Image processing: The nadir-looking MS (R, G, B, NIR) image lines were used, one from each flying height. Two versions of each image were produced using XPro version 4.1:

- No corrections (ASR, at-sensor radiance data)
- Atmospheric (ATM, ground reflectance data)

The ASR-data enables the evaluations of vicarious calibration and the ATM-data is, according to user experiences, the most useful radiometrically corrected product. The BRDF-correction was not tested in this study, because an improved method with water masking will be implemented in the XPro version 4.2.

The images were processed into small patches that covered the ground reference targets (section 2.3, Figure 3). For each target and MS channel the average, minimum, maximum and standard deviation of at-sensor radiances and ground reflectance were calculated in image windows of ground size of 3 m x 3 m.

2.3 In situ measurements

The radiometric reflectance targets included four portable reference reflectance targets (tarps) of the Finnish Geodetic Institute (FGI) (Markelin et al., 2008) installed on the grass football field and various natural and manmade covers (Figure 3, Table 2). During the campaign the nadir spectra of the reference targets were measured using an ASD Field Spec Pro FR spectroradiometer. 12–20 spectra were measured over each target and then averaged to get the final spectra. Before and after each target measurement a reference measurement was made with a calibrated white reference standard (12" Spectralon from Labsphere). The spectra were measured in absolute radiance mode. Afterwards the radiances were scaled with the white reference measurements to produce target reflectances.

Finally, the target reflectances were weighted with the ADS40 channel spectral sensitivities to get the reflectances per MS channel (Figure 4).

The average measurement accuracy of the reference was estimated to be better than 6% for uniform targets and between 6–20% for other targets (Table 2). In the ADS40 data the cross-track viewing angles of the reflectance targets were 3–15˚; the viewing angle in along-track direction was close to zero. The reflectance anisotropy of the reference tarps was analyzed in the laboratory. Tarp P05 with the 10 cm data viewing angles had the highest differences from the nadir reflectance; the difference was 5–10% depending on the channel; differences were 0–5% for other GSDs and targets. Therefore the tarp P05 was not used in the context of the data with the 10 cm GSD.

The Hyytiälä forest research station is equipped with the state-of-the-art SMEAR-II weather station (SMEAR, 2009), which is also part of the NASA AERONET network (Holben et al., 1998). The station provides continuously information about the atmospheric conditions which can be used in radiative transfer calculations.

SH	Target	Time	Sun El.	Sun Az.	Refl.	CV%
A	asphalt	9:59	27.4	120.0	0.140	2.8
B	grass1	10:08	28.4	122.2	0.078	4.9
C	grass2	10:16	29.1	124.3	0.068	9.6
E	sand	10:44	31.7	131.6	0.187	21.0
F	gray gravel1	10:54	32.6	134.4	0.090	8.7
G	gray gravel2	11:00	33.1	136.0	0.090	8.3
H	weeds1	11:09	33.8	138.5	0.100	21.3
I	weeds2	11:14	34.2	139.9	0.062	22.0
P05	tarpaulin 05	10:21	29.6	125.6	0.057	4.9
P20	tarpaulin 20	10:25	30.0	126.6	0.181	2.8
P30	tarpaulin 30	10:29	30.4	127.7	0.261	5.6
P50	tarpaulin 50	10:33	30.8	128.7	0.442	2.9

Table 2. Ground reference targets. SH = short name for target, Time = measurement time (UTC+3), Sun El. = Sun elevation, Sun Az. = Sun azimuth angle (0 = north), Refl. = average target reflectance on green channel (550nm), CV% = ground measurement Coefficient of Variation (100*stdev/mean) for green channel.

Figure 3. Ground reference targets with identifiers in Table 2.

2.4 Validation

2.4.1 The accuracy assessment. The accuracy was assessed by using the ASD reflectance measurements as the reference (absolute accuracy) and by using the imagery with 10 cm GSD as the reference (comparisons of different flying heights). The reflectance differences were normalized with the reference reflectance:

$$Difference = 100*(ATM\_data–reference)/reference\ [\%] \quad (2)$$

This data was calculated for each target, flying height and MS channel. The more detailed analysis was performed using the accurate reference reflectance targets. The root-mean-squared (RMS) values of the differences were calculated for each flying height and channel.

$$RMS = \sqrt{\frac{\sum_{i=1}^{n} Difference^2}{n}}\ [\%] \quad (3)$$

where n is number of targets used in the evaluation. The number of targets was four excluding the data with 10 cm GSD where the P05 was not used due to anisotropy effects (Section 2.3)

2.4.2 Vicarious calibration. The vicarious calibration was performed using simulated at-sensor radiances and ADS40 raw DNs (equation 1) of the four reference reflectance tarps (three tarps for 10 cm data). A linear regression was used to determine the gain and offset parameters. Details of the method are presented in Markelin et. al. (2008). The differences of the gain parameters determined by the vicarious and the laboratory calibration were then calculated. Furthermore, the differences between the ADS40 and simulated at-sensor radiances were calculated using equation (2).

The simulated at-sensor radiances were calculated by the MODO software (ver. 3.0.9, based on a MODTRAN4 ver. 3.1 radiative transfer code) using the ground reflectance measurements and the atmospheric measurements of the SMEAR-II and AERONET stations. The parameters used in the simulations were flying height, solar angles, and the following atmospheric parameters from SMEAR-II: temperature, visibility and CO2, and from AERONET: H2O and O3 (Table 1). A "Midlatitude summer" atmospheric model was used in the simulations. Also the aerosol optical thickness (AOT) at 500 nm from AERONET was observed as an indicator of the general conditions. The effects of aerosol measurement accuracies were not considered in this study.

3. RESULTS AND DISCUSSION

3.1 Validation of XPro-reflectance products

The general view of the reflectance data is shown in Figure 4. The reflectance of grass, asphalt and two reflectance targets (P20 and P50) measured in field and by ADS40 are shown. All measurements showed similar patterns but also differences appeared.

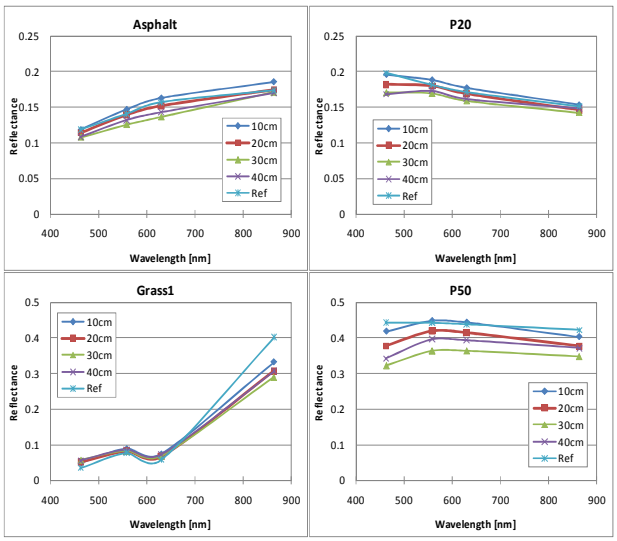


Figure 4. Reflectance of grass and asphalt targets, and two tarps (P20, P50) obtained from atmospherically corrected images and by ASD field measurements (Ref).

3.1.1 Absolute reflectance accuracy. The absolute accuracy was assessed by calculating the differences between the ADS40 reflectance data and the ASD field measurements. Differences (% of reflectance) of individual targets (10 cm and 40 cm data) and RMS-values of differences on reference reflectance tarps are shown in Figure 5a.

The differences on individual objects were in many cases substantially large, varying between 0–70%. The largest differences appeared on spatially non-uniform targets (grass, weeds); for uniform targets (asphalt, sand, reflectance tarps) the differences were less than 20%. Various channels provided different results; the differences were typically the largest in the B channel and the smallest in the R channel. The general performance of the 20 cm and 30 cm GSDs was similar to the 10 cm and 40 cm GSDs. However, the differences appeared to increase with increasing flying height. Furthermore, the performance of data with 30 cm GSD was clearly the worst, because there were some clouds close to the reflectance targets. The above observations indicated that the larger the influences of the atmosphere were, the poorer the absolute accuracy was.

Further analysis was performed using the reflectance tarps. The R, G and NIR channels provided in most cases differences lower than 10%, and in best cases differences lower than 5%

a)

b)

c)

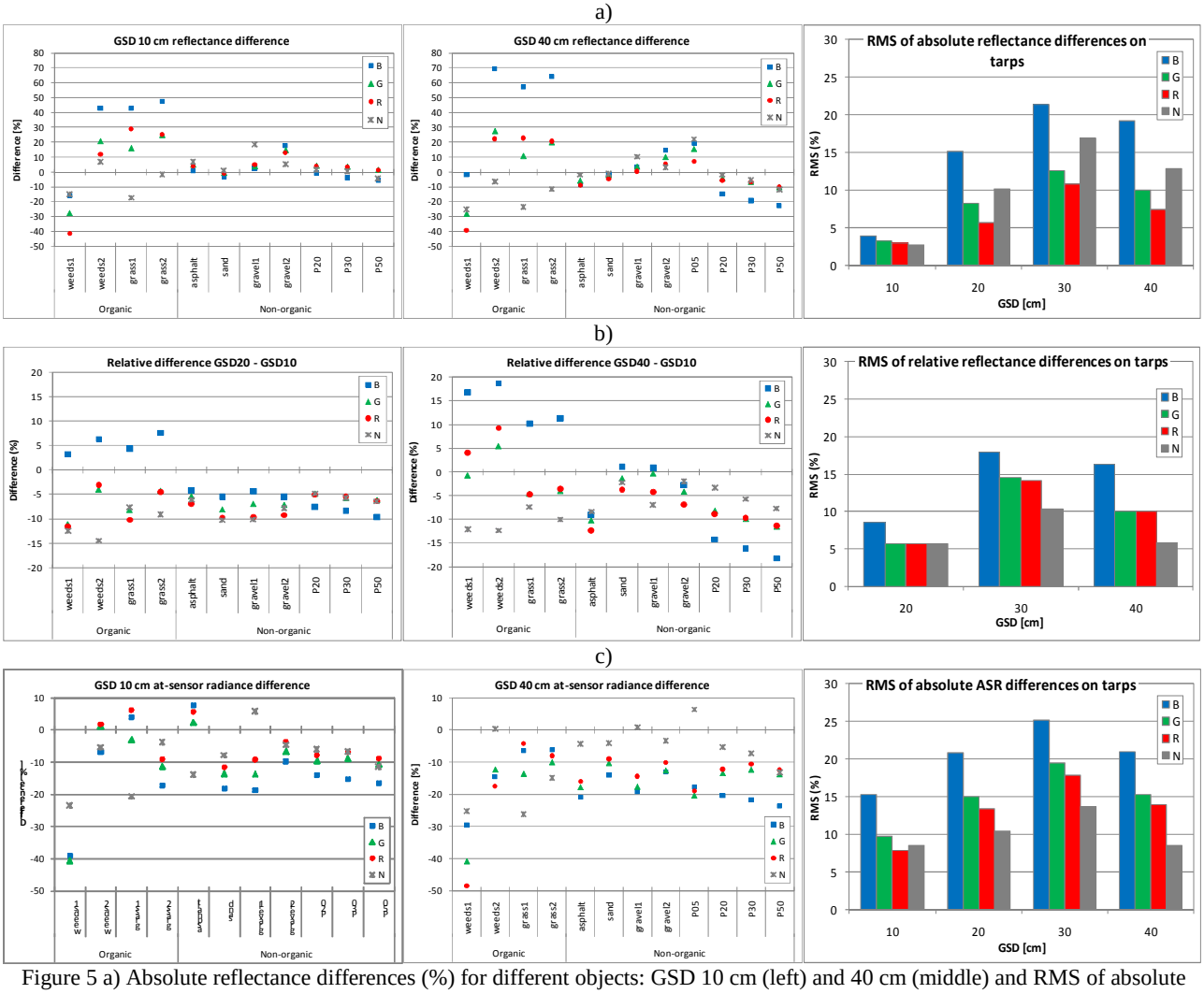


Figure 5 a) Absolute reflectance differences (%) for different objects: GSD 10 cm (left) and 40 cm (middle) and RMS of absolute differences on tarps (right). b) Comparison of different flying heights: GSD 20 cm (left) and GSD 40 cm (middle) and RMS of relative differences on three tarps (right). The image with 10 cm was used as a reference. c) Absolute at-sensor radiance differences (%) for different objects: GSD 10 cm (left) and 40 cm (middle) and RMS of absolute differences on tarps (right).

(Figure 5a). The RMS values of differences are given in Figure 5a (right). With 10 cm data the RMS values were less than 5%. The RMS values of the R channel were 5% or lower on 10 cm and 20 cm data. With 10, 20 and 40 cm data, in most cases, the RMS values were lower than 10%. In the case of the B channel the RMS values were up to 20%.

An unexpected phenomenon was that the differences normalized by the reflectance were dependent on the magnitude of the reflectance. Systematic features in Figure 5a indicated this. As an example, the differences of most uniform objects (tarps, asphalt, sand, gravel) are plotted as a function of the reflectance (NIR-channel) in Figure 6. In this case the systematic features were modeled with a linear regression; the coefficients of determination (R^2) rose up to 0.7. The systematic trend was higher for the NIR and B channels than for the R and G channels, and it increased with the flying height.

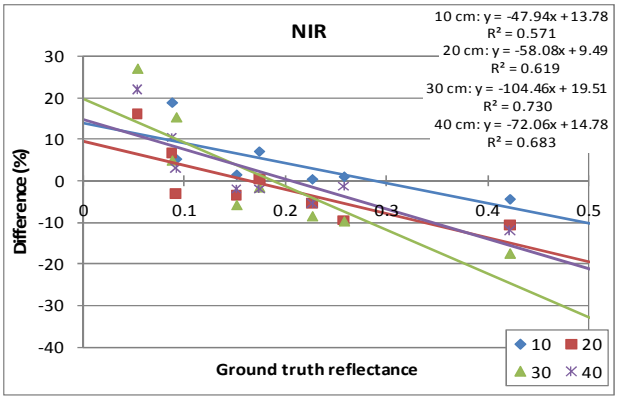


Figure 6. Absolute reflectance difference (%) of the NIR-channel as a function of the reflectance.

3.1.2 Comparisons between different flying heights. The relative accuracy of 20 cm, 30 cm and 40 cm data was assessed by using the 10 cm data as a reference; differences of individual targets and RMS values of differences for the reflectance tarps are given in Figure 5b.

The differences between the flying heights were lower than absolute differences, but similar features appeared in both cases (Section 3.1.1). The relative differences on various objects were 0–20%. The differences on reflectance tarps were mostly less than 10% for images with 20 and 40 cm GSD (excluding B channel) (Figure 5b). The RMS values of the NIR, R and G channels were 5–10% for GSDs 20 cm and 40 cm; RMS values of 30 cm data were 10–18% (Figure 5b, right). The individual channels provided different results. As with the absolute differences, differences were the largest on the B channel.

Again, the differences appeared to be systematically dependent on the magnitude of the reflectance. As in the case of absolute differences, the systematics was the highest on B and NIR channels and 30 cm and 40 cm data.

3.2 Evaluation of vicarious calibration

The differences of ADS40 and simulated at-sensor radiances are shown in Figure 5c. The differences are systematically negative, indicating that the simulation provided too high at-sensor radiance values. The RMS values in Figure 5c (right) are higher than the RMS values of the absolute reflectance differences of XPro ATM-data (Figure 5a (right)), except for the NIR channel; XPro provided thus better results in most cases. Similar, radiance level dependent, systematic differences appeared on the at-sensor radiances derived by MODTRAN4 and the ADS40 as with the relative and absolute analysis of the reflectance data.

The vicarious calibration was compared to the laboratory calibration and differences of gain parameters were calculated. As could be expected based on the systematic difference of the ADS40 and simulated at-sensor radiance (Figure 5c), there was a substantially large difference between the vicarious calibration and laboratory calibration. The differences were the smallest in the R and G channels; for the 10 cm GSD the difference was 10% and for 20 cm and 40 cm GSD the difference was 10–15%. The differences were larger on B and NIR channels and they appeared to increase with the flying altitude. The above results indicated that the simulated at-sensor radiances based on MODTRAN4 radiative transfer code did not perfectly fit the data. However, the weather conditions were not optimal for the vicarious calibration.

3.3 Discussion

The preliminary results of the performance evaluation of the ADS40/XPro at-sensor radiance and atmospherically corrected reflectance products were given here.

The weather conditions were not perfect. The visibility was good, but the clouds changed the diffuse illumination. The conditions were typical of Finland, thus the results are representative in Finland. We feel that it is important to evaluate the performance also in suboptimal conditions.

The flying height influenced the accuracy. The difference between the ADS40 reflectances and the reference measurements increased for denser atmosphere, i.e. for higher flying heights. The results showed differences of 5% or even lower between the ADS40 and the field reference measurements for the 1 km flying height and uniform reference targets. For the 4 km flying height, as low as 10% differences could be obtained. Clouds caused further challenges for the radiometric corrections. These issues could be considered in future developments of the correction software.

Performance varied between the MS channels. The R channel appeared to be the most accurate with 5–7% absolute RMS values in up to 4 km flying height, and the performance of the G channel was very similar. The accuracy of NIR channel was lower than that of the R and G channels (7–12% absolute RMS values) and the B channel was the poorest (up to 20% RMS values). The results suggested a general undercorrection of the atmospheric effects.

The differences between ADS40 and field reflectance measurements were dependent on the target. The tarps, which are spatially uniform targets with low anisotropy, provided lower differences than the structural, natural objects. This can be attributed to the fact that the 1 mrad instantaneous field of view (IFOV) of the ADS40 pixel differs strongly from the 52 mrad IFOV of the ASD spectrometer. BRDF effects from different FOV and different diffuse illumination will result in different reflectances. Also natural vegetated surfaces (e.g. grass and weeds) are difficult to use as reference targets because of strong BRDF effects and changes of moisture during the day. Spatially uniform and temporally stable targets are highly recommended for calibration and validation purposes.

An important observation was that the difference normalized by the reflectance was dependent on the magnitude of the reflectance; more uniform performance would have been expected. This behavior appeared in comparisons of ADS40 and field reference measurements, comparisons between different ADS40 GSDs, and comparisons of MODTRAN4 and XPro derived at-sensor radiances. This result indicated that the atmospheric modeling was not quite accurate in either case to fit the actual atmospheric conditions.

The results from the quantitative performance evaluation of the ADS40 image processing chain could be considered as very good. The processing chain was also very efficient from the operational point of view. The radiometric processing was simple and automated; the user did not have to set any parameters. The corrections were based solely on the image data, thus the processing did not require any field reference targets. For the cases where reflectance targets are available, a recommended improvement for the software would be to enable their use.

The results indicated the importance of the radiometric test field validation. They showed quantitatively the high performance potential of the ADS40 radiometric processing chain and also identified improvement proposals. For the validation process it appeared to be advantageous to apply four calibration targets, distributed on reflectance range of 0.05–0.5.

In our future studies we will further evaluate the performance of the BRDF correction of the XPro software and also evaluate the spectro-directional performance of the system by analyzing the off-nadir views. An important future objective will be to evaluate the performance of the ADS40 imagery in the classification of tree species in Finland.

4. CONCLUSIONS

This article provided preliminary results of the validation of the ADS40 radiometric processing chain. The results indicated that for the current algorithm and the evaluated challenging data set, up to 5% reflectance accuracy could be obtained for uniform targets. The accuracy was influenced by the flying height (1-4 km), channel (R, G, B, NIR), level of cloudiness and target properties. Considering that this was the first, independent quantitative assessment of the absolute accuracy of the ADS40/XPro reflectance products, the results can be considered as very promising. The study also pointed out improvement proposals for the correction method and the validation process.

It can be expected that the future of radiometrically quantitative photogrammetry is bright. On the way towards quantitative radiometric processing chains, the radiometric validations are crucial.

REFERENCES

Beisl, U., 2006. Absolute spectroradiometric calibration of the ADS40 Sensor. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Paris, France, Vol. XXXVI, part 2.

Beisl, U., Telaar, J., and Schönermark, M. V., 2008. Atmospheric correction, reflectance calibration and BRDF correction for ADS40 image data. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Beijing, China, Vol. XXXVII, part B7.

Chavez, P. S., Jr., 1975. Atmospheric, solar, and MTF corrections for ERTS digital imagery. Proc. Am. Soc. Photogrammetry, Fall Technical Meeting, Phoenix, AZ, p. 69.

Chavez, P. S., Jr., 1988. An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. *Remote Sensing of Environment*, 24(3), pp. 459-479.

Fraser, R. S., Ferrare, R. A., Kaufman, Y. J., Markham, B. L., and Mattoo, S., 1992. Algorithm for atmospheric corrections of aircraft and satellite imagery. *International Journal of Remote Sensing*, 13(3), pp. 541-557.

Fricker, P. Raising the bar for multi-band, high-resolution airborne imagery, 2007. In *Photogrammetric Week ´07*; Fritsch, D., Ed.; Wichmann Verlag: Heidelberg, Germany, 2007; pp. 71-79.

Honkavaara, E., Arbiol, R., Markelin, L., Martinez, L., Cramer, M., Bovet, S., Chandelier, L., Ilves, R., Klonus, S., Marshal, P., Schläpfer, D., Tabor, M., Thom, C., and Veje, N., 2009. Digital airborne photogrammetry — A new tool for quantitative remote sensing?—A state-of-the-Art review on radiometric aspects of digital photogrammetric images. *Remote Sensing* 1(3), pp. 577-605.

Holben, B.N., Eck, T.F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J.A., Kaufman, Y.J., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A., 1998. AERONET—A federated instrument network and data archive for aerosol characterization. *Remote Sensing of Environment* 66(1), pp. 1–16.

Markelin, L., Honkavaara, E., Peltoniemi, J., Ahokas, E., Kuittinen, R., Hyyppä, J., Suomalainen, J., and Kukko, A., 2008. Radiometric calibration and characterization of large-format digital photogrammetric sensors in a test field. *Photogrammetric Engineering & Remote Sensing* 74 (12), pp. 1487-1500.

Ryan, R. E., and Pagnutti, M., 2009. Enhanced absolute and relative radiometric calibration for digital aerial cameras. *Proceedings of the Photogrammetric Week 2009*, Stuttgart, Germany.

SMEAR, 2009. Smear research stations. http://www.atm.helsinki.fi/SMEAR/ (Accessed 23 Nov. 2009)

Song, J., Lu, D., and Weseley, M. L., 2003. A simplified atmospheric correction procedure for the normalized difference vegetation index. *Photogrammetric Engineering and Remote Sensing*, 69(5), pp. 521-528.

ACKNOWLEDGEMENTS

The financial support of the Ministry of Agriculture and Forestry of Finland is gratefully acknowledged. We acknowledge all project collaborators, including Estonian Land Board, Leica Geosystems, University of Joensuu, University of Helsinki, and the Finnish Geodetic Institute, and many persons whose hard work made this campaign possible. Especially Timo Tokola, Peep Krusberg, Teemu Hakala, Leena Matikainen and Juha Suomalainen are gratefully acknowledged. We also thank Gerrit de Leeuw for his efforts in establishing and maintaining the Hyytiälä AERONET-site and Janne Levula for providing the SMEAR-II data.

DESCRIBING BUILDINGS BY 3-DIMENSIONAL DETAILS FOUND IN AERIAL PHOTOGRAPHY

P. Meixner *, F. Leberl

Institute for Computer Graphics and Vision, Graz University of Technology, Inffeldgasse 16/II, Graz - (meixner, leberl)@icg.tugraz.at

KEY WORDS: Digital, Interpretation, Modelling, Visualization, Detection, Image, Three-dimensional

ABSTRACT:

A description of Real Properties is of interest in connection with Location-Based Services and urban resource management. The advent of Internet-maps and location aware Web-search inspires the development of such descriptions to be developed automatically and at very little incremental cost from aerial photography and its associated data products. Very important on each real property are its buildings. We describe how one can recognize and reconstruct buildings in 3 dimensions with the purpose of extracting the building size, its footprint, the number of floors, the roof shapes, the number of windows, the existence or absence of balconies. A key to success in this task is the availability of aerial photography at a greater overlap than has been customary in traditional photogrammetry, as well as a Ground Sampling Distance GSD exceeding the traditional values. We use images at a pixel size of 10 cm and with an overlap of 80% in the direction of flight and 60% across the flight direction. Such data support a robust determination of the number of floors and windows. Initial tests with data from the core of the City of Graz (Austria) produced an accuracy of 90% regarding the count of the number of floors and an accuracy of 87% regarding the detection of windows.

1. INTRODUCTION

Urban building models by computer vision have been a topic since the early 1990's (Gruber, 1997). Since 2006, this has evolved into a massive and systematic effort to map buildings in 3D to support a certain location-awareness in Internet-searches. While Google, Yahoo!, Ask and various regional search-providers all implemented 2D systems, Microsoft embarked on a 3D Internet mapping program (Leberl, 2007).
The US website www.zillow.com built an application on top of Microsoft's Internet mapping platform, then denoted as Virtual Earth, now Bing Maps, that attached a description and a value to each property in the USA. Both the description and the value are being taken from public records for property taxes, as shown in Figure 1. Adding the street-side view, one can obtain a rather complete assessment of a property's main characteristics, based on its 2D visualizations from the air and from the street level.

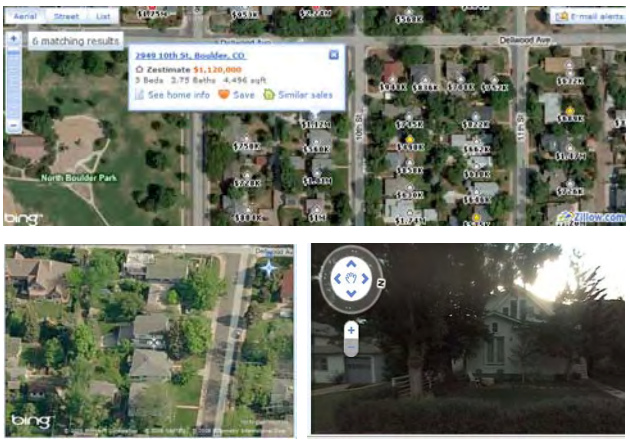


Figure 1. A property valuation website for North America is www.zillow.com. It associates public property tax records with a US$-value, built-out surface area and number of bathrooms to an orthophoto from the Microsoft Bing Maps website (top). The result is an easy access via a known address to the estimated value. In addition, presenting each property also on oblique Microsoft-images (below, left) and accessing the street-view data of Google (below, right) adds considerable visual information per property. However, the image information itself is not entering into the valuation nor description, and there is no searchable data being extracted from the images.

At issue is the development of an ability to describe each property and each building automatically in the absence of detailed and publicly accessible property-tax records. Besides, even if such records exist, they typically will not contain certain details about a property's buildings. Therefore, an ability to describe buildings may be of interest in a broad range of tasks, typically related to the offerings of location-based services. Basing such a description on Internet-public data with its ortho-photos, but augmenting this with data products derived from aerial imagery, would appear to make this description largely a byproduct of aerial mapping, without added cost. Regarding the buildings of a property, major descriptive elements concern its number of floors, roof shape, number of windows, existence of a garage, of a basement or attic, of skylights and chimneys. These elements can be determined automatically, as we will demonstrate in this paper. However, a strictly 2-dimensional data set would be insufficient for the task. We do need 3D data since we approach the building as a 3D structure.
Our approach is based on data that have been created for regular mapping purposes, and we treat such data as input. Using these, we are building specific applications to extract building information. Using a demonstration data set from Graz (Austria) with 216 buildings on 321 parcels, we show that the detection of floors and windows from aerial photography is feasible at a detection rate regarding building floors of 90% and windows of 87%.

* Corresponding author

2. AERIAL PHOTOGRAPHY AND COMPUTED DATA PRODUCTS

Figure 2 is an orthophoto of a segment of the City of Graz and covering 400 m x 400 m. Such orthophotos are today being created from digital aerial photography using pixel sizes of perhaps 10 cm and image overlaps in the range of 80% forward and 60% sideward (Scholz and Gruber, 2009). A point on the ground will thus be imaged 10 times and the orthophoto will not have to have any occluded regions. Both a traditional orthophoto with relief displacements of vertical objects such as buildings and trees is a common product, and increasingly the true orthophoto is as well since the ability to avoid occlusions is essential in this case, and the novel high overlaps ensure that such occlusions get eliminated. However, in order to produce a true orthophoto at a good quality, one needs a Digital Surface Model DSM with well-defined building roof lines to avoid "ragged" building edges. A high-quality DSM requires a 3D capability at an accuracy level that is not needed for traditional orthophotos.

Figure 2. A 400 m x 400 m segment of an orthophoto of the urban core of the city of Graz (Austria). The pixel size is at 10 cm. The orthophoto is of the type "true"; therefore the facades are not visible.

Associated data are computed from the aerial images. They consist firstly of the results of the aerial triangulation with their pose information per image. Given the high overlaps among images and the digital format, the accuracies of the pose and attitude are higher than those of the traditional two image stereo image blocks on film. The demonstration data set in Graz is produced at an accuracy of 10cm on the ground.

Secondly, we have available the DSM plus its filtered Bald Earth DTM (regular rasters). It may be remarkable that the DSM is computed at an interval of the elevation postings at only 2 pixels. Traditional photogrammetry had postulated a distance between elevation postings as a multiple of the height accuracy. That horizontal spacing was recommended to be in the range of perhaps 20 times the elevation error. If one were to assume an elevation error of ± 1 pixel, then the postings were to be spaced 20 pixels apart. However, these recommendations were based on 2-image stereo. This is now changing to a 10-image multi-view geometry (Hartley, Zisserman, 2000), and thus to a concept of "super-resolution", as if the pixel sizes were in effect much smaller than they actually are. The result is a much denser DSM than was ever computed previously (Klaus, 2007). This leads to well-defined horizontal edge information such as building roof lines. This approach also is very competitive with the direct elevation measurements from aerial laser scanners (see Figure 3).

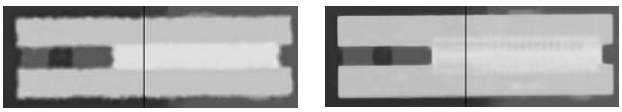


Figure 3. Comparing a building outline obtained from high-overlap digital aerial photography (right) using 8 cm pixels, with the result from an aerial LIDAR measurement (left) using 40 cm postings. This result has been obtained from the Vaihingen test near Stuttgart under supervision by the University of Stuttgart (Cramer and Haala, 2009). This example had been developed in a separate project (Leberl et al, in print).

The third type of derived information is the image classification into roofs, grassy areas, vegetation, water bodies and circulation spaces such as roads, parking spaces, driveways and other impervious surfaces.

3. IDENTIFYING BUILDINGS WITHIN INDIVIDUAL REAL PROPERTIES

3.1 What is a "Building"?

A central task exists to identify "buildings". The definition of a "building" is less obvious than it may initially seem. The imagery needs to be related to parcel maps in the form of cadastral records. Figure 4 presents a cadastral map segment and superimposes it over the orthophoto. The first observation concerns the geometric relationship: the visual data from the imagery are not in complete agreement with the cadastral parcels and a geometric change is needed to achieve a optimum match. The second observation concerns the fact that buildings as seen in aerial imagery cut across property boundaries because they may be attached to one another in dense urban situations.

Figure 4. A cadastral vector data set is superimposed onto the true orthophoto for a segment of the Graz demo site. Note the small discrepancies between the data along property boundary lines manifesting themselves as visual feature in the imagery.

What then is a "building"? In our context, this is a structure of sufficient size within a parcel. Therefore what may be experienced as a single building in aerial photography will be represented by a collection of buildings, each defined by its own parcel. The inverse may also exist, where multiple buildings are defined on a single parcel. This fact leads to a third issue, namely a need to separate smaller structures such as garages or sheds from a building properly.

A fourth topic addresses complex building shapes with many facades. For analysis purposes it would be desirable to have

buildings with only 4 facades. An approach to cope with the complex building shapes may consist of separating an individual building into its parts so that fairly basic building shapes are then be achieved, in analogy to separating the concatenated buildings of urban landscapes along parcel boundaries. In the demo area of Figure 2 we count 216 buildings. Of these, 139 are with a simple rectangular footprint, and at least 2 viewable facades. We find that occlusions form vegetation prevent one often from actually being able to have multiple facades per building available for redundant analyses.
To deal with the second through the fourth issues, we first need to identify the data per parcel.

3.2 Matching Cadastral Parcels with the Orthophoto

In a separate paper we have presented a solution to the problem of mismatches between cadastral and image data (Meixner & Leberl, 2010). Such mismatches can be the result of the different histories of the cadastral data and their focus on 2D local information. We do not allow for a local deformation of the cadastral data. Instead, the cadastral maps are treated as rigid 2D entities where changes are only permitted in rotation and scale. We apply the widely available method of chamfer matching to conflate the vector-type parcel data with the raster-type Orthophoto. Details about the chamfer-matching, the handling of roof overhangs and mismatches between the cadastre and the DSM are illustrated in Meixner and Leberl (2010). This is applicable if sufficient image information is available to define the parcel boundaries by natural features. Major parcel-vector matches with imagery are along street outlines and where fences exist. In our demonstration data set in Graz, we have shown that the initial mismatches in the range of ±7 pixels could be reduced to ±3 pixels.

3.3 Data per Property

Once the orthophoto and cadastral parcel match, one can proceed to cut all data sets along parcel boundaries. Figure 5 illustrates the result for a single property with its DSM, its image classification and its multiple individual overlapping image segments. This example represents a case with no special complexity since there is a single simple building shape with four facades. One complexity is caused by occlusions due to vegetation.

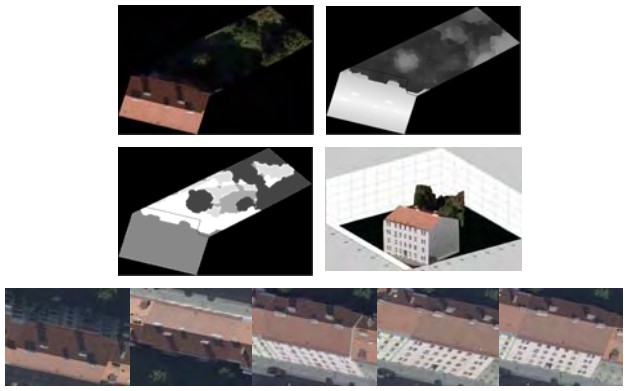


Figure 5. A sample parcel with its data from 10 overlapping aerial photographs, consisting (a) of a True Orthophoto (b) the DSM, (c) the classification layers and (d) a selection of some individual aerial image segments. Also shown is (e) a perspective view of the DSM and the aerial imagery.

4. BUILDING FOOTPRINTS

4.1 Computing Footprints

We have two information sources for building footprints. One is from the image classification of roofs. The other is from the vertical elements of the DSM. The classification typically is based on color and texture, not however, on the 3D information of the DSM. Therefore the two information sources are independent.

(a) Using the Classification Layer “Roofs”
Figure 6 illustrates the classification layer “roof” for a building and its contour in the form of contour pixels. The selection of contour pixels in the binary building layer is trivial. The conversion of the raster- into a vector-format follows a standard procedure according to Douglas D. and Peucker T. (1973). The result consists of straight line segments. Knowledge that this is a building footprint will enter at this point by replacing the line segments by a rectangular shape from a library of such shapes. The match between the line segments and the geometric figure is achieved via a best fit between the geometric shape and the line segments. The measure of fit consists of 4 lines.

Figure 6. The classification layer “building” is based on color and texture. (a) shows the binary layer, (b) its contour in raster and finally in (c) the geometric figure of the footprint.

(b) Using the Vertical Elements of the DSM
A computational pass through the difference between DSM and DTM of a parcel will result in height -postings representing vertical objects.

- Loading the height-postings Hij of the parcel for all rows i and columns j;
- Calculation of the first and second derivative Hij’ and Hij’’ of the height data Hij in each line and column;
- Locating the maximum 1st and 2nd derivatives Hmax’ and Hmax’’ in each line and column, delivering candidate footprint postings;
- For a neighborhood around each candidate footprint location, determine the associated height H of a structure;
- Decide on valid footprint positions from the verticality of the DSM expressed by the values of H’, the curvature expressed by H” and building object expressed by the height H.

The positions of candidate footprint pixels are now in the raster format. We again convert this to line segments as in alternative (a) above. The information now can be fed into the computation of a geometric figure of the building footprint as previously described. This geometric figure is the resulting “building mask”. Other vertical objects may be trees and those also will produce candidate footprint pixels. However, there will not exist straight line segments to replace those pixels and therefore these footprint pixels will get deleted.

4.2 Attaching Heights to the Footprints

The use of the DTM in defining footprints produces, as a by-product, an estimate of an elevation value for each candidate footprint. While this has been computed for candidate positions where a footprint location is possible, this now needs to be converted to a set of elevation values along the path of the footprint. For this purpose the geometric figure of the footprint is placed into the DSM and the elevation profiles get interpolated along the straight lines of the footprint:

For each straight line of the footprint repeat the following process:

- Define positions i,j along the straight footprint line at equal intervals;
- Determine the XY-pixel –locations perpendicular to the line at positions i,j;
- From the short elevation profiles along the pixel locations XY, determine the base height and the top height associated with that footprint element, and thus the elevation difference.

The result of this procedure is a set of elevation profiles along the footprints.

4.3 Buildings Cutting Across Parcel Boundaries

With the elevation values along the footprints, we have the 3D outlines of the buildings. At issue is the situation along a parcel boundary where there may be a valid building footprint, or the building is attached to a structure on the adjacent parcel and the footprint is merely virtual.
To determine whether the footprint is virtual or real, we revisit the elevation data. Along a footprint at the edge of a parcel, one defines a small mask of perhaps 20*20 pixels. If one is dealing with a real footprint, then half of the elevation values should be zero. If the footprint is virtual, then a majority of the elevation values will be large. We select a threshold of 2/3 of all values to be large to determine that the footprint is virtual.

4.4 Small Structures versus Buildings

With elevation profiles along the footprints, we also have the means to separate actual residential housing from detached garages. The latter will have a small surface area of 50 m2 or less and not exceed a height value of 2.5 m.

4.5 Complex Buildings

The split of a complex building into simpler building elements has been discussed by Zebedin et al. (2008) and implemented in a workflow to replace a dense point cloud by simple building geometries.
There exist three measures of complexity for a building. One is the geometric figure of the building's footprint. One may restrict the complexity to be for 4 façades only. The second is the elevation profile along the footprint. One may determine a measure of the building symmetry for the elevations along the footprint: if facades get associated with different building heights, one may have reason to break the building into its parts. The third is the number of local maxima in the elevations of the roof: the roof shape is defined by the elevation values inside the footprint figure. By computing local maxima for those elevations, one will have the means to determine a separation of the building into building elements, each with a separate roof. Zebedin et al. (2008) evaluates the height differences between manual and automatic reconstruction of a building for a test data set of Manhattan (1973 buildings). It shows that 67.51% of the pixels have a height difference smaller than 0.5m, 72.85% differ by less than 1m and 86.91% are within 2m. Details on this method are described in Zebedin et al. (2008).

5. FACADES

The interest is in describing floors and windows, and for this purpose one needs to identify the façades. These are available along the building mask's straight segments, and the elevation profile associated with that line segment. Independent of the actual shape of the façade and where it touches the roof, and how the ground slopes, one can for simplicity define a quadrilateral in 3D space by computing a façade height from the DSM profile. The footprint will define one edge of the quadrilateral in 3D by computing a slope from the DSM values. The end points of the straight line segment define the two opposing vertical edges of the quadrilateral. The DSM-values along the roof line will be replaced by the 4th segment.
Figure 7 illustrates the façade quadrilaterals for the simple building, together with the image texture of one of the aerial photographs covering those facades.

Figure 7. Façades of one building, with the computed quadrilateral for each of the facades. Note that the replacement of the elevation profiles along the building footprints by a straight line serve to obtain a simple façade figure in 3D.

6. FLOORS AND WINDOWS

6.1 Image Texture per Façade

The definition of the façade quadrilaterals produces 4 façade corner points in 3D object coordinates. These must be projected into each of the aerial images to associate image content to each façade. Typically, many aerial images will show the texture of each façade. Figure 8 is an example for one of the separate facades of the building in Figure 7. The projection is based on the pose values of each image from the aerial triangulation.

Figure 8. Of one single façade of the building in Figure 7 one will obtain multiple aerial image segments. These have been rectified into a façade coordinate system. From an aerial image block showing for each object point typically 10 images, not all will contain useful data for a specific vertical façade. Selected here are the 4 best, where "best" is defined as the largest area of a façade quadrilateral in the projection into an image.

6.2 Floors

From the building's appearance, floors get defined by windows. In turn, windows form a defining structure in describing a façade's detail. A procedure for finding a floor count has been developed using the following steps.

For each façade i of a building j, repeat:
Import all n image segments showing this façade i.

- For each image segment repeat:
- Transform the segment into the façade coordinate system.
- Apply a contrast enhancement.
- Apply the Prewitt edge detection horizontally.
- Apply the Prewitt edge detection vertically.
- Convert the maximum horizontal and vertical edge values into a binary format.
- Create for each image row, and image column, a summation of all pixel values, resulting in a vertical and horizontal edge profile.
- From the summation, remove outliers, normalize the values and remove low values as "noise".
- Determine the number of maxima of the sums of vertical gradients and use this as the number of floors.
- Perform a verification by eliminating floors that do not have the proper vertical spacing (minimum distance between floors); and removal values from along the edges of the image texture inside the façade quadrilateral.

This approach will result in data as illustrated in Figure 9.

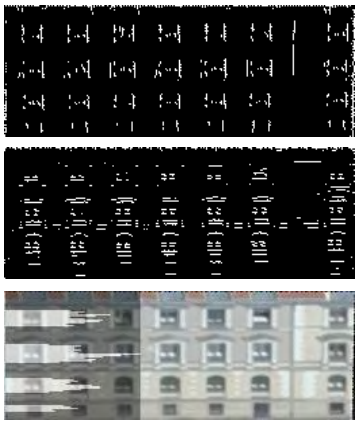


Figure 9. Binary Prewitt edges in (a) are vertical, in (b) horizontal. The sums of edge values are shown in (c) as a count of the number of floors.

A floor count can be applied to each of a set of overlapping façade images. If there were a discrepancy in the result, some logic would have to get applied to resolve the ambiguity.

6.3 Windows

Window detection has been of some interest in recent years. Algorithms like "boosting" have been applied by Nguyen et al. (2007) to detect cars and windows in aerial images. Cech and Sara (2007) have developed a window detection based on a library of window shapes. Lee and Nevatia (2004) have based their approach on edge images. These approaches have been subjected to only limited experimental analysis, but are generally reported to find windows in a rather robust manner.

Given our floor counts, we are reusing the intermediate Prewitt edges to also find the windows. An approach that simply "intersects" the locations along the pixel rows and columns with the maximum edge sums will work if all windows are regularly arranged. While this is often the case, it is not always true. Therefore Lee and Nevatia (2004) have proposed a variation of the approach.
To refine the locations of the windows a one dimensional search for the four sides of a window is performed. For every line of a window hypothesized lines are generated by moving the lines to its perpendicular direction. The refined positions of the windows are determined where the hypothesized line has the best score for the window boundary. For a more detailed description of the used algorithm read Lee and Nevatia (2004).
The big advantage of this method is that one can also use images with lower resolution, and that not only rectangular windows but almost all window designs can be automatically detected rather quickly without training the program in advance.
The window count is applicable in each image segment of a given façade, separately. Or one might want to merge the edge data sets and apply a single window detection to the sum of all edges. Initial tests have shown that the window count is a rather robust method that delivers no discrepancies between the separate images of one façade in the examples chosen thus far.
A comparison of the various different methods for window detection should be performed and will be the subject of ongoing work.

6.4 Multiple Facades per Building

The redundancy not only applies to the image coverage per façade from the high overlaps of aerial photography. We also find that we have multiple measures for the number of floors from multiple facades. These must be consistent with one another. It is possible that a building has different floor counts on a sloping terrain. Since the "bald Earth" as well as the slope of a building footprint are known, they must enter into the floor count.
Figure 7 presented facades of one building. Figure 10 illustrates the floor counts and detected windows in each façade of that one building. As one can easily determine, the automated floor count and the count of the windows is consistent with a visual inspection.

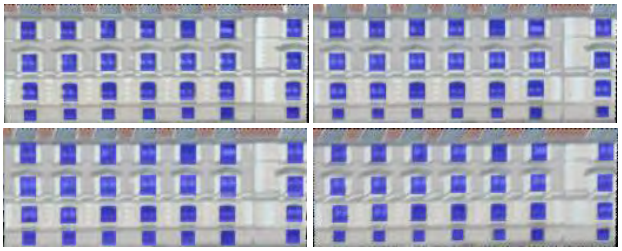


Figure 10. Four facades of one building from Fig. 7 lead to independent floor counts and window counts. It hss to be noted that the floor counts and the number of windows coincide with the visual inspection.

We have extended this exercise to a selection of 150 properties in the Graz demo data set. In those properties we have identified 102 buildings with a total of 225 facades. The total number of floors was 387, the number of all windows was 2646. Running the approach through this data set results in the following:

Success rate of Building detection: 100%, all 102 building were found.

Success rate of Floor detection: 90% of the 387 floor were correctly counted.

Success rate of Window detection: 87.1% of the 2646 windows were correctly counted.

7. CONCULSION: TOWARDS AN EXTENSIVE PROPERTY DESCRIPTION

The search for a description of individual buildings per property is but an element in a larger effort. The development of as detailed a description of real properties will have the buildings as the most important element, but other features of a property are also in need of a description. One will want to consider the land, the vegetation, the impervious surfaces, even the interaction between properties casting shadows or affecting privacy. And one will also be interested in the traffic, distances to businesses or public transportation etc. A full system for property descriptions will involve business addresses, traffic information, street network information, as well as sun angles.

In the current contribution we have focused on basic descriptions of buildings. This involves the definition of a building on a property, even if two buildings are connected along a property line. It deals with complex buildings having many facades and a complex roof-scape. From the outside, thus from aerial imagery, one can count the floors and windows, and identify the window areas on a façade for further analysis. At this stage of research we are beginning with the experimental evaluation of the various approaches. We will have to cope with occlusions from vegetation, with ambiguities regarding garages and sheds, the difficulties arising from an inability of matching parcel maps with aerial imagery, and with ambiguities from basement and attic windows.

Initial results are encouraging. Using 150 properties with 102 buildings having 387 facades and 2646 windows, 90% of all floors and 87.1% of all windows were found automatically. The result addresses, however, a specific situation in a mature core area of Graz (Austria). Reasons for misclassifications regarding floors and windows result from inaccuracies of the DTM, occlusions from vegetation and other buildings, partial shadows on the facades, very complex facades and steep camera angles. All these reasons for misclassifications have to be analyzed very carefully. Fore that the building interpretation has to be repeated by increasing the sample data in one city, and then by looking at vastly different environments such as a coastal resort environments, historical small towns, alpine terrains and industrial zones.

REFERENCES

Cech J., R. Šara , 2007. Windowpane detection based on maximum a-posteriori labeling. Technical Report TR-CMP-2007-10, Center for Machine Perception, K13133 FEE Czech Technical University, Prague, Czech Republic.

Cramer, M. & Haala, N., 2009. *DGPF project: Evaluation of digital photogrammetric aerial based imaging systems - overview and results from the pilot centre.* Published at ISPRS Workshop High-Resolution Earth Imaging for Geospatial Information, Hannover, Germany, June 2 - 5. Digitally published on CD, 8 pages.

Douglas D., T. Peucker, 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature, *The Canadian Cartographer* pp. 112-122.

Gruber M., 1997. *„Ein System zur umfassenden Erstellung und Nutzung dreidimensionaler Stadtmodelle“*, Dissertation, Graz University of Technology,2007.

Hartley R, A. Zisserman, 2000. Multiple View Geometry for Computer Vision. *Cambridge University Press*, 1st Edition.

Klaus A., 2007. *Object Reconstruction from Image Sequences.* Dissertation, Graz University of Technology, 2007.

Leberl F., 2007. Die automatische Photogrammtrie für das Microsoft Virtual Earth *Internationale Geodätische Woche Obergurgl.* Chesi/Weinold(Hrsg.), Wichmann-Heidelberg-Publishers, pp. 200-208

Leberl F., A. Irschara, T. Pock, P. Meixner, M. Gruber, S. Scholz, A. Wiechert (in print) Point Clouds from Laser Scanning Versus 3d Vision. *Photogrammetric Engineering and Remote Sensing.*

Lee S.C., R. Nevatia, 2004. Extraction and Integration of Window in a 3D Building Model from Ground View Images. *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition* CVP’04

Meixner P, F. Leberl, 2010. From Aerial Images to a Description of Real Properties: A Framework. Manuscript submitted for publication, Graz University of Technology – Institute of Computer Graphics and Vision, Graz.

Nguyen T., H. Grabner, B. Gruber, H. Bischof, 2007. On-line Boosting for Car Detection from Aerial Images. *Proceedings of the IEEE International Conference on Research, Innovation and Vision for the Future* (RIVF'07), pages 87-95.

Scholz S., M. Gruber, 2009. Radiometric and Geometric Quality Aspects of the Large Format Aerial Camera UltraCam Xp. *Proceedings of the ISPRS*, Hannover Workshop 2009 on High-Resolution Earth Imaging for Geospatial Information, XXXVIII-1-4-7/W5, ISSN 1682-1777

Zebedin L., Bauer J. Karner K., Bischof H., 2008. Fusion of Feature- and Area-Based Information for Urban Buildings Modeling from Aerial Imagery. *Proceedings of the ECCV* 2008, Marseille, France, pages 873-886

TERRAIN ECHO PROBABILITY ASSIGNMENT BASED ON FULL-WAVEFORM AIRBORNE LASER SCANNING OBSERVABLES

Werner Mücke[a], Christian Briese[b], Markus Hollaus[a]

[a]Institute of Photogrammetry and Remote Sensing, Vienna University of Technology
[b]Christian Doppler Laboratory "Spatial Data from Laser Scanning and Remote Sensing"
Gußhausstraße 27-29, 1040 Vienna, Austria
wm,cb,mh@ipf.tuwien.ac.at
http://www.ipf.tuwien.ac.at

KEY WORDS: Vegetation, LIDAR, analysis, classification, laser scanning, estimation

ABSTRACT:

Airborne laser scanning (ALS) has become a widely used method for data acquisition in various fields of engineering over the past few years. The latest generation of commercially available ALS systems, the so-called full-waveform ALS systems, are capable of detecting the whole backscattered waveform, which needs to be analysed in post-processing in order to detect the individual echoes. During this signal processing step additional observables, such as the amplitude and the width of the backscattered echo, are derived. The hereby produced 3D point cloud holds additional information about the radiometric and geometric characteristics of the objects within the footprint area of the laser beam. In this paper point cloud samples of different ground cover are examined regarding their distribution of amplitude and echo width. Subsequently, a method for employing these observables for the assignment of probabilities, whether an echo is more likely to stem from terrain or not, is presented. These probabilities can also be interpreted as individual weights that are assigned to the single points and can be used in subsequent digital terrain modelling (DTM) algorithms for a derivation of more accurate DTMs.

1 INTRODUCTION

The sampling of the earth's surface with laser technology for obtaining information about its geometric structure has become an efficient and wide-spread method for data acquisition over the past few years. 3D point clouds from airborne laser scanning (ALS), which is also referred to as airborne LiDAR (light detection and ranging), provide a data basis for various applications and have been used in different fields of engineering, such as forestry (Hyyppä et al., 2008; Naesset, 2007), urban monitoring (Dorninger and Pfeifer, 2008; Höfle et al., 2009), hydrology (Mandlburger et al., 2009; Briese et al., 2009; Höfle et al., 2009) or archaeology (Doneus et al., 2008).

Most of the above mentioned applications have in common that they rely on an accurate digital terrain model (DTM), derived on the basis of the point cloud. The quality of the DTM itself is, among other influencing factors (Kraus et al., 2004), dependent on the reliability to eliminate off-terrain echoes (Kraus et al., 2004; Karel et al., 2006). Conventional methods for classifying the point cloud into terrain and off-terrain echoes, a process also called filtering, employ various geometric criteria. This might be the distance to prior computed surfaces (Axelsson, 2000; Kraus and Pfeifer, 1998; Pfeifer et al., 2001), relations of planimetric distance and height difference (Vosselman, 2000) or normal vectors as a homogeneity criteria in a segmentation based approach (Tóvari and Pfeifer, 2005). However, reflections from near terrain objects, e.g. lower under storey, cannot be distinguished by geometric criteria alone. Especially near ground vegetation poses two problems. The first problem concerns the range resolution. If the vegetation is very low, the range difference between two consecutive targets may become too short for the detector to separate them. Consequently, only one target is identified, which features a measured distance that results from an overlap of the two actual reflections. The resulting point is then located somewhere in between them (Kraus, 2007). Secondly, areas with dense vegetation feature only little to no penetration at all. This is crucial if the trend of the surface changes significantly below the impenetrable vegetation and no echoes from the terrain are detected. In both cases, echoes tend to be wrongly classified as ground points. Consequently, a DTM surface computed on the basis of a point cloud including such off-terrain echoes, might run through the lowest vegetation levels and therefore above the actual terrain. As these errors can be in the range of several decimetres, they are critical for DTM based application where high accuracy is required (Doneus et al., 2008).

Currently, two different types of ALS sensors, which can be distinguished by their method of echo detection, are commercially available. The so-called discrete recording systems are able to record the range and amplitude of one or more consecutive discrete echoes. In contrast, the so-called full-waveform (FWF) digitizers, are capable of detecting and storing the whole emitted and backscattered signal. To then obtain the individual echoes, the recorded waveform has to be reconstructed in post-processing and a decomposition algorithm, which can be individually adapted, has to be applied. Recent papers describe different methods for ALS waveform analysis and echo detection (Wagner et al., 2006; Roncat et al., 2008; Mallet et al., 2009). During the process, the echoes are detected and the range of the scanner to the target, as well as additional variables are derived. In addition to the amplitude, the width of the backscattered signal, also commonly known as the echo width, is determinable.

The usage of these additional observables opens up new prospects for DTM generation from ALS data, although very rarely used so far. Wagner et al. (2008) stated that the width of the backscattered echo is dependent on the vertical distribution of small surface elements within the footprint area of the laser beam. The canopy, under storey or near ground vegetation are assumed to have larger variations in vertical directions and consequently larger echo widths than the terrain. Based on this fact, Doneus et al. (2008) used an empirically derived echo width threshold, pre-classifying presumable off-terrain echoes in the input point cloud for the hierarchic robust filtering (Pfeifer et al., 2001). In Lin and Mills (2009), a point labelling process, determining terrain points us-

ing a threshold for the echo width is applied to complement the individual 3D points. This additional surface information is integrated in a DTM generation approach employing Axelsson's progressive densification method (Axelsson, 2000).

However, applying hard thresholds on datasets poses several difficulties. On the one hand, the derived thresholds are always sensor specific and do not necessarily apply for others. On the other hand, the above mentioned strategies tend to eliminate a certain number of points based on a-priori determined thresholds. This implies the possibility of creating false negatives, meaning excluding reflections that might very well stem from terrain. These echoes are permanently lost for subsequent filtering steps.

The method described in this paper considers these limitations and disadvantages of pre-classifying echoes based on a fixed echo width threshold. Rather than using thresholds on either one of the FWF-observables (amplitude and echo width), it focuses on the modelling of the distribution of the echo widths dependent on amplitude values. As for the derivation of a DTM only the last echoes are relevant, probabilities indicating whether they are more likely to be a terrain echo or not are assigned to the echoes. These probabilities can be interpreted as individual weights and can be used as a-priori weights in existing filtering algorithms. Hence, the whole point cloud is preserved and augmented with additional information, which can subsequently be used for a more accurate derivation of DTMs.

In the following section 2 the study area is described. Section 3 deals with the theoretical background of the proposed method (see section 3.1), a description of the point cloud analysis (see section 3.2) and the probability assignment (see section 3.3). Finally, the results are summarized and discussed in section 4 and a conclusion is given in section 5.

2 STUDY AREA AND DATA

An ALS data set collected over the city of Eisenstadt, capitol of county Burgenland, Austria, was used in this paper. As study area a small sample within the Schlosspark Eisenstadt was created. The ALS data were acquired under leaf-off conditions in March 2007. A RIEGL LMS-Q560 laser scanner, which is equipped with a full-waveform recorder, was employed. The main technical specifications can be found on the distributors website (Riegl, 2009a). The scanner was carried by a fixed-wing aircraft as well as a helicopter alternatively, which operated at an average flying altitude of 600 m, the scan angle was set to $\pm 22.5^{\circ}$ and the average distance of the single trajectories was 90 m. This resulted in a large overlap of the ALS strips and, consequently, rather high point density of 18 echoes per m^2 for the whole data set. Using Gaussian decomposition, as described in Wagner et al. (2006), the single echoes were extracted from the raw waveform data and a 3D point cloud was obtained. For adequate geo-referencing the method proposed by Kager (2004) was applied. This process allows to reduce discrepancies between overlapping ALS strips. The produced high quality 3D point cloud was then used to derive a digital surface model (DSM) utilizing moving planes interpolation. Furthermore, a DTM using hierarchic robust interpolation (Pfeifer et al., 2001) was generated. Both methods are implemented and documented in the software package SCOP++ (SCOP++, 2008). The DTM was used to compute the normalized heights of the single echoes for point selection and later verification purposes. A hill-shading of the study area can be seen in figure 1a. The obtained point cloud consisted of 57.3% single echoes (only one reflection in the shot), 27.8% shots with two, 11.8% with three 2.7% with four and 0.4% with more than four consecutive target reflections.

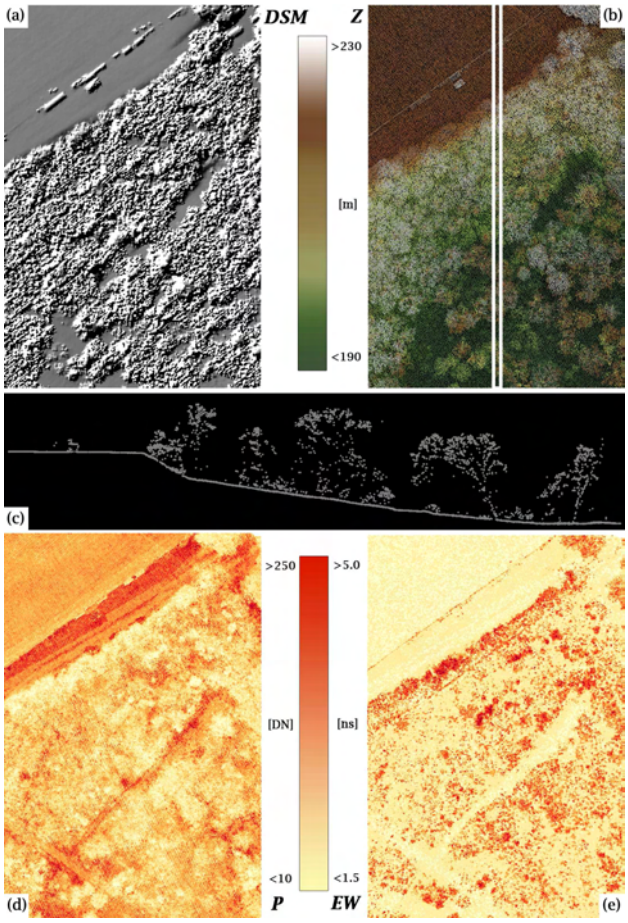


Figure 1: ALS study area in Schlosspark Eisenstadt (120 x 170 m). (a) DSM (grid size 0.25 x 0.25 m) created of all echoes; (b) height-coded point cloud showing all echoes (Z), white bars indicate location of profile; (c) profile; (d)amplitudes (P) of the last echoes; (e) echo widths (EW) of the last echoes.

3 METHOD

3.1 Theoretical Background

As written in section 1 and based on Wagner et al. (2008), the echo width provides information about the height distribution of small scatterers within the footprint area of the laser beam. A planar area, perpendicular to the laser beam direction, features no height variation at all. It is therefore assumed that a laser beam impinging approximately rectangular on such terrain should cause a backscattered signal with an echo width equal or similar to the width of the emitted pulse (Wagner et al., 2004). According to its specifications, the RIEGL LMS-Q560 laser scanner emits pulses with a duration of 4 ns, describing the full width of the pulse at its half maximum (FWHM). It has to be stated that due to the utilized software for echo extraction, the derived echo widths do not represent the FWHM, as it is common practice in signal processing and commercial software (e.g. Riegl (2009b)), but the standard deviation of the fit Gaussian curve. The width of the emitted pulse has to be divided by a constant factor of

$$2 * \sqrt{2 * ln2} = 2.3548 \quad (1)$$

in order to correspond to the echo widths derived by the applied echo extraction method, which results in a value of 1.6986 ns (Mücke, 2008).
Examination of the echo width image in figure 1e supports this theory. In the north-western corner a football court is located,

which is a flat non-tilted plane featuring also the lowest echo widths of about 1.75 ns in the sample area. The pathway through the forest (centre of the image), which is not overgrown by trees, also shows reflections of comparable widths. The largest echo widths in the area stem from vegetation echoes, characterized by values well above 2 ns and up to 5 ns. Also the lower amplitude values can be found in the vegetated area (see figure 1d). This is probably the result of a potentially higher number of consecutive echoes in the penetrable canopy and multi-layered vegetation below. The more reflections per shot, the less energy can be scattered back by the individual targets and consequently the last echoes feature the lower amplitude values (Wagner et al., 2008).

3.2 Point cloud analysis

To confirm our assumption, single echoes from open terrain areas were selected and scatter plots showing their amplitudes and respective echo widths were produced (see figure 2). Compared to the width of the emitted pulse (dotted blue line in figure 2), the four samples, taken from reflections over gravel (1452 points), sand (1059 points), asphalt (1427 points) and grassland (4628 points), seem to correspond quite well. Except for a minimal positive shift with respect to the width of the emitted pulse and increased scattering of the echo widths with decreasing amplitude values. This scattering was also observed by Wagner et al. (2006), explained as a characteristic of FWF decomposition using a Gaussian model. They stated that the method is robust for strong echoes and becomes less reliable for weak ones. Subsequently, we examined if the amplitudes and widths of single and last echoes from forest terrain behave similar as the ones from open terrain. Figure 3 shows a scatter plot produced from reflections of an overgrown area. They were selected using the previously calculated normalized heights (dZ) with respect to the DTM and a dZ $<=$ 0.2 m threshold (see section 2), so they approximate the terrain as good as possible. We found that the 0.2 m threshold was adequate given the definition uncertainty of a forest terrain covered by fallen leafs, tree roots and creepers of any kind. Further reducing the threshold did not significantly minimize the resulting terrain point cloud, whereas an increase to 0.5 m had the opposite effect. By visual comparison we examined that lots of points representing relevant near terrain vegetation were included.

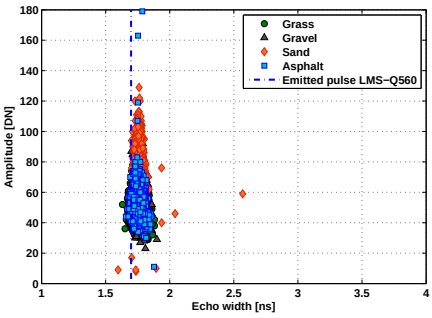


Figure 2: Scatter plot of selected single echoes from open terrain (grass, gravel, sand, asphalt); width of the emitted pulse is shown by dotted blue line.

3.3 Probability Assignment

We expect the scatter plot in figure 3 to depict the distribution for amplitudes and echo widths of terrain echoes in vegetated areas and base the probability assignment on this assumption. First, a distribution function dependent on the amplitudes and respective echo widths is created. It is fit to the distribution on the left side

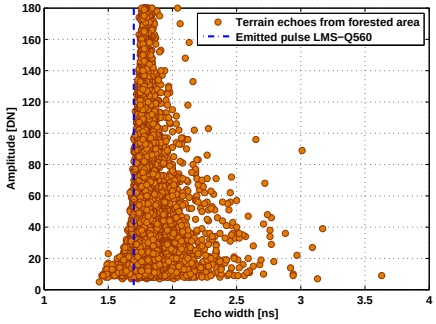


Figure 3: Last echoes from forest terrain selected by height thresholding (dZ $<=$ 0.2 m); width of the emitted pulse is shown by the dotted blue line.

of the median of the echo widths, because there the detection accuracy we can expect from the applied Gaussian decomposition is demonstrated. We then flip the function vertically along the median (see figure 4), so the two functions combined now outline the range of amplitude and echo width values for reflections that stem from terrain with a high probability. 3D points within this area are given the highest weights, which is defined as 95% (green zone in figure 4). As we do not expect a shortening of the echo widths due to vegetation but rather the opposite, every echo left of the area outlined by the distribution function (green zone) is given the lowest weight of 5%. In order to avoid hard thresholds in the transition from high to low weight, we defined a buffer zone along the flipped curve. Inside of it a linear function is used for applying the individual weights, ranging from 95% to 5%. Every point outside the transition zone is considered not to represent terrain and is therefore given the lowest weight (5%). This method tends to give the highest weights to points belonging to terrain with a high probability, points assumed to be off-terrain echoes gain the lowest weights.

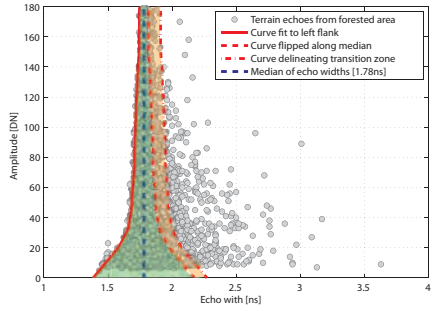


Figure 4: Last echoes from forest terrain; fit distribution function (red lines); highest weights w = 95% (green), transition zone 95% $<$ w $<$ 5% (orange), lowest weights w = 5% are assigned outside the green and orange zones.

4 RESULTS AND DISCUSSION

The distribution of the forest terrain echoes (figure 3) shows similar characteristics as the one from the open terrain echoes (figure 2). The scattering of the echo widths also increases with decreasing amplitudes. However, it features echoes with very low amplitudes, which are not present in the single echoes from open terrain and can be explained by the persistent loss of energy due to the detection of consecutive targets in the vegetated area. Also the echo widths cover a wider range up to 3 ns, indicating that some of the selected terrain points represent rough surfaces, e.g.

points on tree roots or stems. Apart from that a slight shift compared to the emitted pulse is recognizable in all of the produced point cloud samples. At this time the cause of this shift is uncertain. Probably the incidence angle of the laser shot might influence the echo width, causing it to become wider with small (acute) angles. Essentially, this corresponds to a widening of the reflected signal caused by height variation. However, the examined echoes stemming from open areas represent near horizontal terrain (e.g. left part of profile in 1c representing the football court) and are located in the middle of the strip, roughly below the trajectory of the flight. So the angle of deflection is small and consequently the incidence angles are rather obtuse, a change of the echo widths caused by this is therefore unlikely. Another reason could be that the width of the emitted pulse is not constantly 4 ns, but sometimes wider, subsequently causing bigger echo widths in the backscattered signals. But to our knowledge this has not been explored so far.

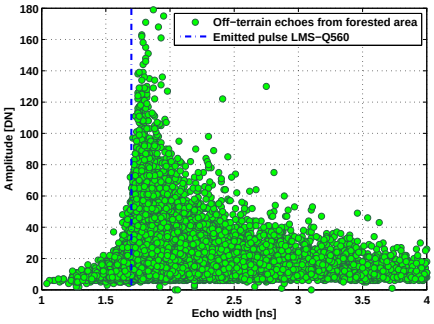


Figure 5: Off-terrain echoes from a forested area selected by height thresholding (last echoes dZ $>$ 0.2 m); width of the emitted pulse is shown by dotted blue line.

Further, the distribution function derived in section 3 (figure 4) was compared to the scatter plot of the selected off-terrain last echoes in the forested area (see figure 5), which were generated by height thresholding (dZ $>$ 0.2 m) as well. Although there is a large overlap, the distribution is very asymmetric and shifted towards higher echo widths, featuring a significant amount of echoes we can expect to be given the lowest weight.

Consequently, the distribution function was used to assign weights to the single and last echoes within the study area, because these are usually the lowest points and therefore considered to represent terrain. For a validation of the results the afore calculated normalized heights were employed. Initially we selected terrain points in the forested area by applying a threshold of 0.2 m (dZ $<=$ 0.2 m) (see section 3.2) on the normalized heights of the single and last echoes. We repeated that for our study area (see figure 1), producing a terrain point cloud that consisted of 184868 points, starting from a total of 226767 points. Comparing this number to the number of points that were assigned weights of w = 95%, we found that 92% of the terrain echoes selected by height thresholding were given the maximum weight (w $=$ 95%). Additionally, we extracted the echoes that gained the maximum weight and those that gained less (w $<$ 95%) and produced a histogram of the normalized heights (see figure 6). It clearly reveals that more than 80% of the echoes with maximum weight are located below 0.5 m. Apart from that, 7.1% of the points with weights of less than 95% (points that are unlikely to stem from terrain) are also in that height range. More detailed investigations and visual examinations have shown that these echoes stem from lower vegetation or tree stems, therefore also being correctly weighted. This explanation is supported by figure 7b, where the low weighted echoes (w $<$ 95%) color-coded by their respective normalized height dZ are shown. The points are within the dark green area feature normalized heights of 0.5 m and below.

The remaining part of the single and last echoes, featuring high as well as low weights, are spread over the entire range of normalized heights occurring in our study area. This points out a limitation of the probability assignment, which is partly rooted in the applied echo detection method. The lower the amplitude, the higher the tolerance for points gaining maximum weight. Strong first echoes in the canopy might have large echo widths and be therefore correctly weighted. However, for consecutive echoes less energy is left (Wagner et al., 2008), and thus, as we have already pointed out in section 1, the echo width estimation becomes less trustworthy. This probably results in reflections from branches and stems high above the terrain having smaller echo widths, although they represent rather rough surfaces. For this reason they are wrongly given high weights. Figure 7a shows

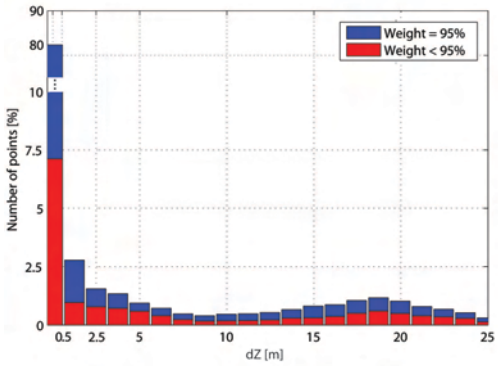


Figure 6: Histogram of single and last echoes from sample area. Points that were weighted with 95% (blue) and less than 95% (red).

the points with a weight of w = 95%. The white square outlines echoes in the canopy (dZ $>$ 20 m). The apparent explanation for these points being given the maximum weight is that the dense canopy acts as a single scatterer, featuring too little height variation within the footprint of the laserscanner to be detected. Consequently, they have similar characteristics as terrain echoes. The same applies for thick branches of trees or stems with large diameters, which can also be seen as extended targets (Jelalian, 1992), meaning they are bigger than the footprint size. However,

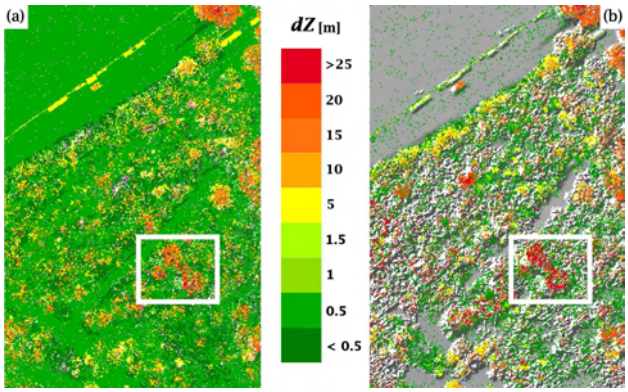


Figure 7: (a) DSM overlain with single and last echoes featuring maximum weights w = 95% (b) DSM overlayed with single and last echoes featuring weights w $<$ 95%. The white square shows the area with dense canopy.

these points do not pose a problem for conventional filtering algorithms, as they are well elevated and can be reliably detected and

eliminated by the geometric analysis of the local neighbourhood. As pointed out in section 1, it is rather the near ground vegetation that causes troubles in DTM generation and the proposed classification method has proven to be valuable in the detection of reflections from such objects.

5 CONCLUSIONS AND FUTURE WORK

Our investigations have demonstrated that the full-waveform observables amplitude and echo width have potential for the assignment of probabilities whether an ALS point represents terrain or not. The suggested method managed to label 92% of the terrain echoes correctly (see section 4) and additionally correctly detected reflections from near ground vegetation. However, we found that the approach can not produce exclusive terrain point clouds, given the fact that amplitude and echo width are metrics that do not discriminate sufficiently to do so. A reliable determination of off-terrain points without the utilization of geometric criteria seems rather difficult. But as presented in other papers (Briese et al., 2007; Lin and Mills, 2009), conventional filtering methods profit from a-priori seperated vegetation echoes which could be found with the help of full-waveform observables. The approach proposed in this paper produces a point cloud enriched by individual weights for each point, which can now be used in subsequent filtering steps. In our further work we will concentrate on the integration of the individual weights into the hierarchic robust interpolation. Apart from that we plan on acquiring a proper reference dataset with terrestrial laser scanning which we will use for the derivation of a very detailed and highly accurate DTM to compare it to the ALS DTM and adequatetly validate the employed filtering method.

ACKNOWLEDGEMENTS

The authors would like to thank Michael Doneus, Department for Prehistory and Early History of the University of Vienna, Austria, for providing the used ALS data set. Further we would like to thank Andreas Roncat for fruitful discussions. This study was partly funded by the TransEcoNet project implemented through the CENTRAL EUROPE Program co-financed by the ERDF.

References

Axelsson, P., 2000. DEM generation from laser scanner data using adaptive TIN models. In: International Archives of Photogrammetry and Remote Sensing, XXXIII, B4, Amsterdam, Netherlands, pp. 111–118.

Briese, C., Doneus, M., Pfeifer, N. and Melzer, T., 2007. Verbesserte DGM-Erstellung mittels full-waveform airborne Laserscanning. In: 3-Ländertagung DGPF, SGPBF, OVG, Basel.

Briese, C., Mandlburger, G. and Ressl, C., 2009. Automatic break line determination for the generation of a dtm along the river main. In: ISPRS Workshop Laserscanning 2009, Paris, FRANCE.

Doneus, M., Briese, C., Fera, M. and Janner, M., 2008. Archaeological prospection of forested areas using full-waveform airborne laser scanning. Journal of Archaeological Science 35(4), pp. 882–893.

Dorninger, P. and Pfeifer, N., 2008. A comprehensive automated 3d approach for building extraction, reconstruction, and regularization from airborne laser scanning point clouds. Sensors 8(11), pp. 7323–7343.

Höfle, B., Mücke, W., Dutter, M., Rutzinger, M. and Dorninger, P., 2009. Detection of building regions using airborne lidar - a new combination of raster and point cloud based gis methods. In: Geospatial Crossroads at GI-Forum 09 – Proceedings of the Geoinformatics Forum Salzburg.

Höfle, B., Vetter, M., Pfeifer, N., Mandlburger, G. and Stötter, J., 2009. Water surface mapping from airborne laser scanning using signal intensity and elevation data. Earth Surface Processes and Landforms 34(12), pp. 1635–1649.

Hyyppä, J., Hyyppä, H., Leckie, D., Gougeon, F., Yu, X. and Maltamo, M., 2008. Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. International Journal of Remote Sensing 29(5), pp. 1339–1366.

Jelalian, A. V., 1992. Laser Radar Systems. Artech House, Boston.

Kager, H., 2004. Discrepancies between overlapping laser scanning strips - simultaneous fitting of aerial laser scanner strips. In: International Archives of Photogrammetry and Remote Sensing, XXXV, B/1, Istanbul, Turkey, pp. 555–560.

Karel, W., Briese, C. and Pfeifer, N., 2006. Dtm quality assessment. In: International Archives of Photogrammetry and Remote Sensing, XXXVI, 2, Vienna, Austria.

Kraus, K., 2007. Photogrammetry – Geometry from Images and Laser Scans. 2 edn, de Gruyter.

Kraus, K. and Pfeifer, N., 1998. Determination of terrain models in wooded areas with airborne laser scanner data. ISPRS Journal of Photogrammetry and Remote Sensing 53, pp. 193–203.

Kraus, K., Briese, C., Attwenger, M. and Pfeifer, N., 2004. Quality measures for digital terrain models. In: International Archives of Photogrammetry and Remote Sensing, XXXV, B2, Istanbul, Turkey, pp. 113–118.

Lin, Y. and Mills, J., 2009. Integration of full-waveform information into the airborne laser scanning data filtering process. In: Laser Scanning 2009, Paris.

Mallet, C., Lafarge, F., Bretar, F., Roux, M., Soergel, U. and Heipke, C., 2009. A stochastic approach for modelling airborne lidar waveforms. In: F. Bretar, M. Pierrot-Deseilligny and G. Vosselman (eds), Laserscanning 2009 – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII, Part 3/W8, pp. 201–206.

Mandlburger, G., Hauer, C., Höfle, B., Habersack, H. and Pfeifer, N., 2009. Optimisation of lidar derived terrain models for river flow modelling. Hydrology and Earth System Sciences 13(8), pp. 1453–1466.

Mücke, W., 2008. Analysis of full-waveform airborne laser scanning data for the improvement of DTM generation. Master's thesis, Institute of Photogrammetry and Remote Sensing, Vienna University of Technology.

Naesset, E., 2007. Airborne laser scanning as a method in operational forest inventory: Status of accuracy assessments accomplished in scandinavia. Scandinavian Journal of Forest Research 22(5), pp. 433–442.

Pfeifer, N., Stadler, P. and Briese, C., 2001. Derivation of digital terrain models in the SCOP++ environment. In: Proceedings of OEEPE Workshop on Airborne Laserscanning and Interferometric SAR for Detailed Digital Terrain Models, Stockholm, Sweden.

Riegl, 2009a. www.riegl.com. Homepage of the company RIEGL Laser Measurement Systems GmbH, accessed: October 2009.

Riegl, 2009b. www.riegl.com. Datasheet of the commercial full-waveform analyzing software RiAnalyze by RIEGL GmbH, accessed: November 2009.

Roncat, A., Wagner, W., Melzer, T. and Ullrich, A., 2008. Echo detection and localization in full-waveform airborne laser scanner data using the averaged square difference function estimator. The Photogrammetric Journal of Finland 21(1), pp. 62–75.

SCOP++, 2008. Institute of photogrammetry and remote sensing. `www.ipf.tuwien.ac.at/products/products.html`.

Tóvari, D. and Pfeifer, N., 2005. Segmentation based robust interpolation - a new approach to laser data filtering. In: International Archives of Photogrammetry and Remote Sensing, XXXVI, 3/W19, Enschede, The Netherlands, pp. 79–84.

Vosselman, G., 2000. Slope based filtering of laser altimetry data. In: International Archives of Photogrammetry and Remote Sensing, XXXIII, B3, Amsterdam, Netherlands, pp. 935–942.

Wagner, W., Hollaus, M., Briese, C. and Ducic, V., 2008. 3d vegetation mapping using small-footprint full-waveform airborne laser scanners. International Journal of Remote Sensing Vol. 29, No. 5, pp. 1433–1452.

Wagner, W., Ullrich, A., Ducic, V., Melzer, T. and Studnicka, N., 2006. Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner. ISPRS Journal of Photogrammetry and Remote Sensing 60(2), pp. 100–112.

Wagner, W., Ullrich, A., Melzer, T., Briese, C. and Kraus, K., 2004. From single-pulse to full-waveform airborne laser scanners: potential and practical challenges. In: International Archives of Photogrammetry and Remote Sensing, XXXV, Istanbul, Turkey.

AIRBORNE DOPPLER NAVIGATION SYSTEM APPLICATION FOR MEASUREMENT OF THE WATER SURFACE BACKSCATTERING SIGNATURE

A. Nekrasov [a]

[a] Taganrog Institute of Technology, Southern Federal University, 347922 Taganrog, Russia
Interdepartmental Center of Research in Environmental Sciences, University of Bologna, 48123 Ravenna, Italy
alexei-nekrassov@mail.ru

KEY WORDS: Sea, Measurement, Navigation, Active, Algorithms, Radar, Systems

ABSTRACT:

A method for measuring the water surface backscattering signature using the airborne Doppler navigation system in addition to its standard navigational application is discussed. A case of an airplane circle flight measurement of the azimuth normalized radar cross section curve of the water surface is considered. This is done in the range of middle incidence angles by the Doppler navigation system. The system operates in the scatterometer mode and uses a fore-beam directed to the right side at a typical mounting angle in the vertical plane that is not so far from nadir at a straight flight. An algorithm for measuring the water surface backscattering signature is proposed.

1. INTRODUCTION

Many researchers have been investigating the microwave backscattering signatures of the sea and ocean surfaces (Carswell et al, 1994; Chelton and McCabe, 1985; Feindt et al, 1986; Hildebrand, 1994; Masuko et al, 1986; Melnik, 1980; Moore and Fung, 1979; Wismann, 1989). However, the mechanics of interactions between water surfaces and microwaves have not been well studied in detail.

The typical method for describing sea clutter is in the form of the normalized radar cross section (NRCS), the statistical distribution of the NRCS, the amplitude correlation and the spectral shape of the Doppler returns.

To describe the radar backscatter from the water surfaces, three major scattering models are used: the Kirchhoff or Physical Optics model, the composite-surface or two-scale model, and the Bragg model. The Kirchhoff model assumes a perfectly conducting surface (unless it is modified to include the Fresnel reflection coefficient) and applies from small to intermediate incidence angles without shadowing effects. Apart from the implicit dependence on the Fresnel coefficient, there is no polarization dependence. The two-scale model assumes that the radar backscatter arises from a large number of slightly rough ripples, distributed over the long ocean waves. It has polarization dependence. These two models are generally used to interpret the data acquired by the synthetic aperture radar and real aperture radar of a variety of sea/oceanic features, including swell waves and internal waves. The Bragg model applies only to the slightly rough surfaces under low wind conditions (it is often used to describe the scattering from ripples in the two-scale model). The Bragg model has been used to interpret the ocean currents by high-frequency Doppler radar measurements at large incidence angles (Ouchi, 2000).

To explain adequately the microwave scattering signature of the water surface and to apply its features to remote sensing, a set of experiments, namely, experimental verification of the combined frequency, azimuth and incidence angles, and wind speed variations of the NRCS are required (Masuko et al, 1986). For that study, a scatterometer, radar designed for measuring the surface scatter characteristics, is used.

Research on microwave backscatter by the water surface has shown that the use of a scatterometer also allows an estimation of near-surface wind speed and direction because the NRCS of the water surface depends on the wind speeds and directions. Based on experimental data and scattering theory, a significant number of empirical and theoretical backscatter models and algorithms for estimation of a near-surface wind vector from satellite and airplane has been developed (Long et al, 1996). It is also very important for safe landing of amphibian aircrafts on the water surface.

To study a microwave backscattering signature of the water surface from airplane, an airborne scatterometer is used. The measurements are typically performed at either a circle track flight using fixed fan-beam antenna or a rectilinear track flight using rotating antenna (Carswell et al, 1994; Masuko et al, 1986; Wismann, 1989). Unfortunately, a microwave narrow-beam antenna has considerable size at Ku-, X- and C-bands that makes its placing on a flying apparatus difficult. Therefore, a better way needs to be found.

At least two ways can be proposed. The first way is to apply the airborne scatterometers with wide-beam antennas as it can lead to the reduction in the antenna size. The second way is to use the modified conventional navigation instruments of a flying apparatus in a scatterometer mode that seems more preferable.

From that point of view, a promising navigation instrument is the Doppler navigation system (DNS). Previous research of the DNS has shown that it allows measurement of the wind vector over the water surface when it operates as a multi-beam scatterometer under the horizontal rectilinear flight of a flying apparatus (Nekrasov, 2005a; Nekrasov, 2005b). Now, a method to measure the water surface backscattering signature by the airborne DNS operating in the scatterometer mode at aircraft circle flight in addition to its standard navigation application is discussed in this paper.

2. DOPPLER NAVIGATION SYSTEM

DNS is the self-contained radar system that utilizes the Doppler effect (Doppler radar) for measuring the ground speed and drift angle of flying apparatus and accomplishes its dead-reckoning navigation (Sosnovskiy and Khaymovich, 1987).

The internationally authorized frequency band of 13.25 to 13.4 GHz has been allocated for airborne Doppler navigation radar. A center frequency of 13.325 GHz of the band corresponds to a wave length of 2.25 cm. This frequency represents a good compromise between too low a frequency, resulting in low-velocity sensitivity and large aircraft antenna size and beam widths, and too high a frequency, resulting in excessive absorption and backscattering effects of the atmosphere and precipitation. (Earlier Doppler radars operated in two somewhat lower frequency bands, i.e., centered at 8.8 and 9.8 GHz, respectively, but now these bands are no longer used for stand-alone Doppler radars.) (Kayton and Fried, 1997).

Measurement of the wind vector and drift angle of flying apparatus is based on change of a Doppler frequency of the signal reflected from the underlying surface, depending on a spatial position of an antenna beam. Usually, an antenna of the DNS has three beams (λ-configuration; beams 1, 2, and 3) or four beams (x-configuration; beams 1, 2, 3, and 4) located in space as represented in Figure 1. An effective antenna beamwidth is of 3° to 10° (Kolchinskiy et al, 1975). Power reasons (DNS should operates over water as well as over land) and sensitivity of the DNS to velocity influence a choice of a mounting angle of a beam axis in the vertical plane θ_0.

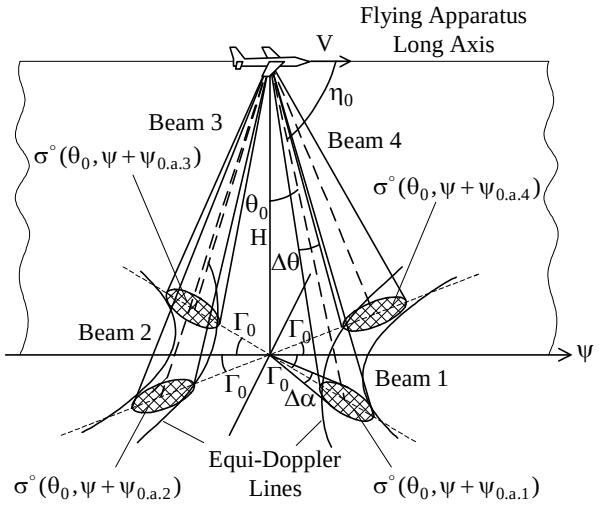


Figure 1. Typical spatial location of the DNS beams: λ-configuration (beams 1, 2, and 3) and x-configuration (beams 1, 2, 3, and 4)

Figure 2 shows curves of the NRCS of underlying surface versus incidence angle for radar system operating in the frequency band (Ke-band) currently assigned to Doppler navigation radar (Kayton and Fried, 1997). It is seen from the curves that for most types of terrain the NRCS decreases slowly with increase of the beam incidence angle. However, for water surfaces, the NRCS falls radically as the incidence angle increases and assumes different values for different conditions of sea state or water roughness. For the typical Doppler-radar incidence angles of 15° to 30° (Kolchinskiy et al, 1975), the NRCS is considerably smaller for most sea states than for land and decreases markedly for the smoother sea state. Therefore, a conservative Doppler-radar design is based on an NRCS for the smoothest sea state over which the aircraft is expected to navigate. (Very smooth sea states are relatively rare).

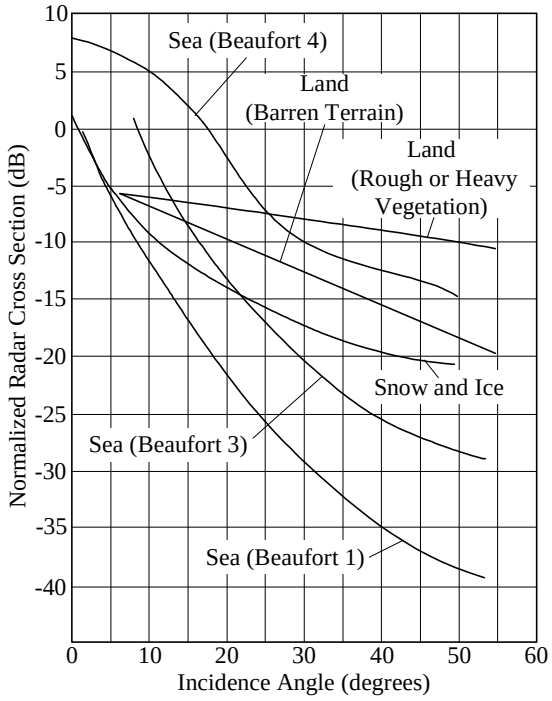


Figure 2. Backscattering cross section per unit surface area (NRCS) versus incidence angle for different terrains at Ke-band (Kayton and Fried, 1997)

There are two basic antenna system concepts used for drift angle measurement. These are the fixed-antenna system, which is used in most modern systems, and the track-stabilized (roll-and-pitch-stabilized) antenna system. For physically roll-and-pitch-stabilized antenna systems, the value of an incidence angle remains essentially constant and equal to the chosen design value. For fixed-antenna system, a conservative design is based on the NRCS and range for the largest incidence angle that would be expected for the largest combination of pitch and roll angles of the aircraft (Kayton and Fried, 1997).

The choice of a mounting angle of a beam axis in the inclined plane η_0 (nominal angle between antenna longitudinal axis and central beam direction) represents a compromise between high sensitivity to velocity and over-water accuracy, which increases with smaller mounting angles of a beam axis in the inclined plane, and high signal return over water, which increases for larger mounting angles of a beam axis in the inclined plane. Most equipments use a mounting angle of a beam axis in the inclined plane of somewhere between 65° and 80° (Kayton and Fried, 1997). The choice of a mounting angle of a beam axis in the horizontal plane Γ_0 depends on the desired sensitivity to drift, which tends to increase with increasing that mounting angle. For the typical Doppler-radar, mounting angles of a beam axis in the horizontal plane are of 15° to 45° (Kolchinskiy et al, 1975).

The relationship among those mounting angles is (Kayton and Fried, 1997)

$$\cos\eta_0 = \cos\Gamma_0 \cos\theta_0 \quad (1)$$

The mounting angle of a beam axis in the horizontal plane should satisfy the following condition $\Gamma_0 > \beta_{dr.\max}$, where $\beta_{dr.\max}$ is the maximum possible drift angle (Sosnovskiy and Khaymovich, 1987). The mounting angle of a beam axis in the inclined plane is defined by requirements to the width of a Doppler spectrum of the reflected signal Δf_D, which depends on the effective antenna beamwidth in the inclined plane $\theta_{a.incl}$; $\theta_{a.incl} \approx 5^\circ$ for DNS. The relative width of a Doppler spectrum $\Delta f_D / F_D$ is given by (Davydov et al, 1977)

$$\frac{\Delta f_D}{F_D} = \frac{\theta_{a.incl}}{\sqrt{2}} \tan\eta_0 \quad (2)$$

where F_D = Doppler frequency, $F_D = \frac{2V_g}{\lambda}\cos\eta_0$

V_g = aircraft velocity relative to the ground

λ = radar wavelength

To perform high accuracy measurements with the DNS, the following condition should be provided (Davydov et al, 1977)

$$\frac{\Delta f_D}{F_D} \le 0.1 \div 0.2 \quad (3)$$

Thus, from (2) and (3), the mounting angle of a beam axis in the inclined plane should satisfy the following condition

$$\eta_0 \le \arctan\left[(0.1 \div 0.2)\frac{\sqrt{2}}{\theta_{a.incl}}\right] \quad (4)$$

From (4), assuming that the effective antenna beamwidth in the inclined plane is typical and equal to 5°, the condition of choice the mounting angle of a beam axis in the inclined plane is

$$\eta_0 \le 58.3^\circ \div 72.2^\circ \quad (5)$$

Then, using (1), the areas of admissible mounting angles of beam axes could be obtained. Lower limits corresponding to the maximum admissible mounting angles of beam axis in the inclined plane and area of typical mounting angles of beam axes in the vertical and horizontal planes are represented in Figure 3. Trace 1 and trace 2 are the lower limits corresponding to the maximum admissible mounting angles of beam axis in the inclined plane of 58.3° (lower limit of high accuracy of measurement at $\Delta f_D / F_D = 0.1$) and 72.9° (lower limit of sufficient high accuracy of measurement at $\Delta f_D / F_D = 0.2$), respectively. A dash line displays the area of typical mounting angles of beam axes in the vertical and horizontal planes.

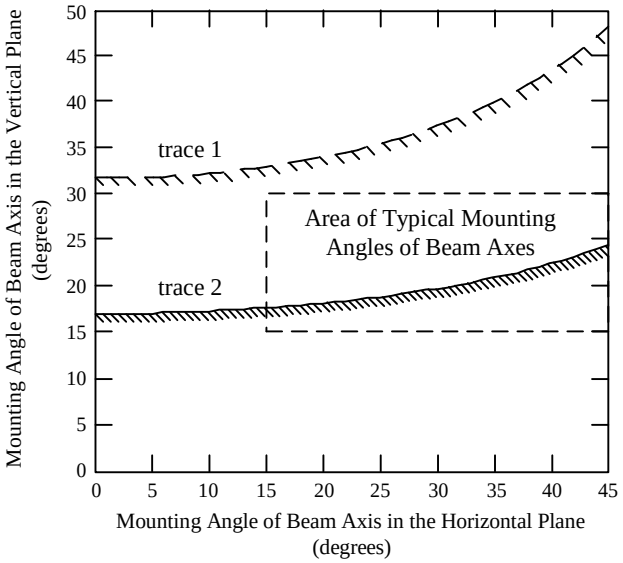


Figure 3. Lower limits corresponding to the maximum admissible mounting angles of beam axis in the inclined plane and area of typical mounting angles of beam axes in the vertical and horizontal planes: trace 1 is the lower limit corresponding to the maximum admissible mounting angle of beam axis in the inclined plane of 58.3° (lower limit of high accuracy of measurement at $\Delta f_D / F_D = 0.1$); trace 2 is the lower limit, which corresponds to the maximum admissible mounting angle of beam axis in the inclined plane of 72.9° (lower limit of sufficient high accuracy of measurement at $\Delta f_D / F_D = 0.2$); dash line is the contour of the area of typical mounting angles of beam axes in the vertical and horizontal planes

Figure 3 demonstrates that for typical mounting angles of beam axes, sufficient high accuracy of measurement by the DNS is provided for the most part of the area of typical mounting angles in the vertical and horizontal planes. The measurement accuracy rises with increase of the beam incidence angle in the vertical plane.

The DNS multi-beam antenna allows selecting a power backscattered by the underlying surface from different directions, namely from directions corresponding to the appropriate beam relative to the aircraft course ψ, e.g. $\psi_{0.a.1}$, $\psi_{0.a.2}$, $\psi_{0.a.3}$, and $\psi_{0.a.4}$ in Figure 1. Each beam provides angular resolutions in the azimuthal and vertical planes, $\Delta\alpha$ and $\Delta\theta$ respectively.

3. WATER SURFACE BACKSCATTERING SIGNATURE MEASUREMENT

As the azimuth NRCS curve can be obtained using the circle track flight for a scatterometer with an inclined one-beam fixed-position antenna (Masuko et al, 1986), one beam of the DNS operating in the scatterometer mode can be used.

Let the flying apparatus make a horizontal rectilinear flight with the speed V at some altitude H above the mean sea surface, and the DNS has a roll-and-pitch-stabilized antenna system. Then, the NRCS values obtained with beams 1, 2, 3, 4 would be $\sigma^\circ(\theta_0, \psi + \psi_{0.a.1})$, $\sigma^\circ(\theta_0, \psi + \psi_{0.a.2})$, $\sigma^\circ(\theta_0, \psi + \psi_{0.a.3})$, and $\sigma^\circ(\theta_0, \psi + \psi_{0.a.4})$ respectively.

Let the beam 1 be used to measure the water surface backscattering signature because both λ- and x-configured DNS have it. As a beam 1 axis is directed to the right side and its mounting angle in the vertical plane is not so far from nadir (at a straight flight), the circle flight with the left roll should be completed (Figure 4).

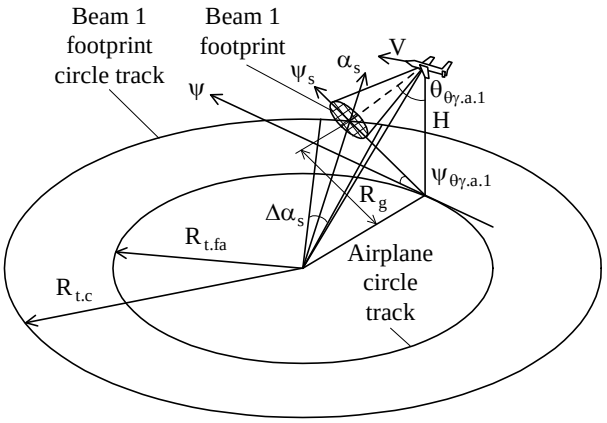


Figure 4. Circle flight geometry for measurement of the water surface backscattering signature

The DNS uses a fixed-antenna system (physically non-stabilized to the local horizontal), the flying apparatus makes the circle flight, and so the values of the incidence angle of the beam and the beam location in azimuthal plane are not equal to the chosen design values. An actual incidence angle of the beam 1 $\theta_{\theta\gamma.a.1}$ and its actual azimuth direction $\psi_{\theta\gamma.a.1}$ relative to the aircraft current course (aircraft ground track) are as following

$$\theta_{\theta\gamma.a.1} = \arctan\left(\sqrt{\tan^2(\arctan(\tan\theta_0 \sin\psi_{0.a.1}) + \gamma_{fa}) + \rightarrow} \rightarrow \overline{\tan^2(\arctan(\tan\theta_0 \cos\psi_{0.a.N}) + \theta_{fa})}\right) \quad (6)$$

$$\psi_{\theta\gamma.a.1} = \begin{cases} \arctan\left(\dfrac{\tan(\arctan(\tan\theta_0 \sin\psi_{0.a.1}) + \gamma_{fa})}{\tan(\arctan(\tan\theta_0 \cos\psi_{0.a.1}) + \theta_{fa})}\right) \\ \textit{for } \tan(\arctan(\tan\theta_0 \cos\psi_{0.a.1}) + \theta_{fa}) \geq 0 \\ \pi + \\ \arctan\left(\dfrac{\tan(\arctan(\tan\theta_0 \sin\psi_{0.a.1}) + \gamma_{fa})}{\tan(\arctan(\tan\theta_0 \cos\psi_{0.a.1}) + \theta_{fa})}\right) \\ \textit{for } \tan(\arctan(\tan\theta_0 \cos\psi_{0.a.1}) + \theta_{fa}) < 0 \end{cases} \quad (7)$$

where $\psi_{0.a.1}$ = azimuthal mounting angle of the beam 1 axis relative to the aircraft course ψ, $\psi_{0.a.1} = \Gamma_0$
γ_{fa} = roll angle of flying apparatus (right roll is positive)
θ_{fa} = pitch angle of flying apparatus (pull-up is positive)

Then, the current NRCS value obtained with the beam 1 is $\sigma^\circ(\theta_{\theta\gamma.a.1}, \psi + \psi_{\theta\gamma.a.1})$. The radius of the flying apparatus turn $R_{t.fa}$, the ground range R_g, and the radius of turn of the selected cell middle point $R_{t.c}$ are described by the following expressions obtained using the geometry of Figure 4

$$R_{t.fa} = \frac{V^2}{g \tan\gamma_{fa}} \quad (8)$$

$$R_g = \frac{H}{\tan\theta_{\theta\gamma.a.1}} \quad (9)$$

$$R_{t.c} = \sqrt{R_{t.fa}^2 + R_g^2 + 2R_{t.fa}R_g \sin\psi_{\theta\gamma.a.1}} \quad (10)$$

where g = acceleration of gravity, $g = 9.81$ m/s^2

The time of the airplane turn for 360° (360-degree turn) T_{360° is given by (Mamayev et al, 2002)

$$T_{360^\circ} = \frac{2\pi V}{g \tan\gamma_{fa}} \quad (11)$$

Usually, the 360-degree azimuth space is divided into 72 or 36 sectors under the circle NRCS measurement. The azimuth size of a sector observed is 5° or 10°, respectively. A middle azimuth of the sector is the azimuth of the sector observed. The azimuth size of a sector relative to the center point of circle of the airplane track is $\Delta\alpha_s$, and the middle azimuth of a sector is α_s. The NRCS samples obtained from the sector and averaged over all measurement values in that sector give the NRCS value $\sigma^\circ(\theta_{\theta\gamma.a.1}, \psi_s)$ corresponding to the real observation azimuth angle of the sector ψ_s that is

$$\psi_s = \psi_{\psi_s} + \psi_{\theta\gamma.a.1} \pm 360^\circ \quad (12)$$

where ψ_{ψ_s} = flying apparatus course corresponding to the real observation azimuth angle of the sector

Real observation azimuth angles of the sector beginning $\psi_{s.b}$ and of the sector ending $\psi_{s.e}$ are

$$\psi_{s.b} = \psi_s + \Delta\alpha_s / 2 \pm 360^\circ \quad (13)$$

$$\psi_{s.e} = \psi_s - \Delta\alpha_s / 2 \pm 360^\circ \quad (14)$$

The time of a sector view T_s and the number of samples N_s that can be obtained from the sector are represented by the following expressions

$$T_s = T_{360^\circ} \frac{\Delta\alpha_s}{360^\circ} \quad (15)$$

$$N_s = \frac{T_s V}{0.5a} \quad (16)$$

where a = antenna length in the direction of flight.

Thus, to obtain an azimuth NRCS curve of the water surface at middle incidence angles under flying apparatus circle flight by the DNS operating in the scatterometer mode and using a fore-beam directed to the right side at a typical mounting angle in the vertical plane that is not so far from nadir at a straight flight, the measurement should be perform in accordance with a scheme of Figure 5.

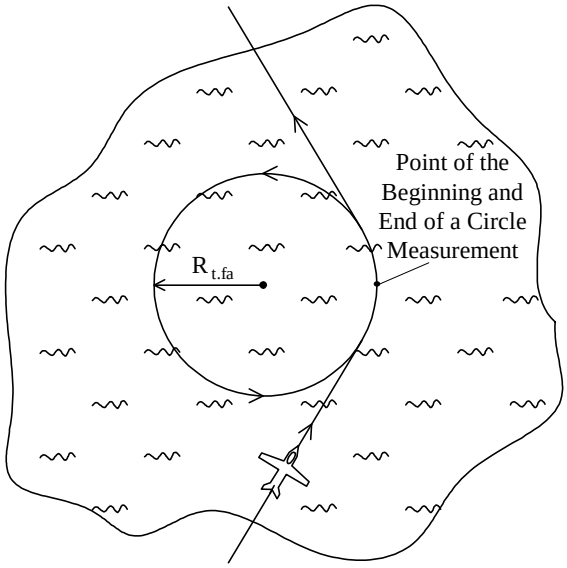


Figure 5. Scheme of a circle flight for measurement of the water surface backscattering signature

Measurement is started when a stable flight at the given altitude, speed of flight, roll and pitch has been established. Measurement is finished when the azimuth of the measurement beginning is reached. To obtain a greater number of NRCS samples for each sector observed several consecutive full circle turns for 360° may be done.

4. CONCLUSIONS

The study has shown that the airborne DNS operating in the scatterometer mode can be used for measuring the water surface backscattering signature in addition to its typical navigation application.

As the azimuth NRCS curve for a scatterometer with an inclined one-beam fixed-position antenna can be obtained using the circle track flight, one beam of the DNS can be used.

Since the mounting angle of the beam axis in the vertical plane is located not so far from nadir (at a straight flight), the circle flight with a small roll should be carried out to provide the azimuth NRCS curve measurement in the range of middle incidence angles.

The algorithm and method proposed in the paper can be used for the DNS enhancement, for designing an airborne radar system for operational measurement of the sea roughness characteristics. They are particularly important for ensuring safe landing of amphibian aircraft on the water surface, for example under search and rescue missions or fire fighting in the coastal areas and fire risk regions.

REFERENCES

Carswell, J.R.; Carson, S.C.; McIntosh, R.E.; Li, F.K.; Neumann, G.; McLaughlin, D.J.; Wilkerson, J.C; Black, P.G. and Nghiem, S.V., 1994. Airborne scatterometers: Investigating ocean backscatter under low- and high-wind conditions. *Proc. IEEE*, 82(12), pp. 1835-1860.

Chelton, D.B. and McCabe, P.J., 1985. A review of satellite altimeter measurement of sea surface wind speed: With a proposed new algorithm. *J. Geophys. Res.*, 90(C3), pp. 4707-4720.

Davydov, P.S.; Zhavoronkov, V.P.; Kashcheyev, G.V.; Krinitsyn, V.V.; Uvarov, V.S. and Khresin, I.N., 1977. *Radar Systems of Flying Apparatuses*. Transport, Moscow, USSR, 352 p., in Russian.

Feindt, F.; Wismann, V.; Alpers W. and Keller, W.C., 1986. Airborne measurements of the ocean radar cross section at 5.3 GHz as a function of wind speed. *Radio Science*, 21(5), pp. 845-856.

Hildebrand, P.H., 1994. Estimation of sea-surface wind using backscatter cross-section measurements from airborne research weather radar. *IEEE Trans. Geosci. Remote Sens.*, 32(1) pp. 110-117.

Kayton, M. and Fried, W.R., 1997. *Avionics Navigation Systems*. John Wieley & Sons, New York, USA, 773 p.

Kolchinskiy, V.Ye.; Mandurovskiy, I.A. and Konstantinovskiy, M.I., 1975. *Autonomous Doppler Facilities and Systems for Navigation of Flying Apparatuses*. Sovetskoye Radio, Moscow, USSR, 432 p., in Russian.

Long, D.G.; Donelan, M.A.; Freilich, M.H.; Graber, H.C.; Masuko, H.; Pierson, W.J.; Plant, W.J.; Weissman, D. and Wentz, F., 1996. Current progress in Ku-band model functions. Tech. Rep. MERS 96-002, Brigham Young Univ., USA, 88 p.

Mamayev, V.Ya.; Sinyakov, A.N.; Petrov, K.K. and Gorbunov, D.A., 2002. *Air navigation and elements of navigation calculations*. GUAP, Saint Petersburg, Russia, 256 p., in Russian.

Masuko, H.; Okamoto, K.; Shimada M. and Niwa, S., 1986. Measurement of microwave backscattering signatures of the ocean surface using X band and Ka band airborne scatterometers. *J. Geophys. Res.*, 91(C11), pp. 13065-13083.

Melnik, Yu.A., 1980. *Radar Methods of the Earth Exploration*. Sovetskoye Radio, Moscow, USSR, 264 p., in Russian.

Moore, R.K. and Fung, A.K., 1979. Radar determination of winds at sea. *Proc. IEEE*, 67(11), pp. 1504-1521.

Nekrasov, A., 2005a. On possibility to measure the sea surface wind vector by the Doppler navigation system of flying apparatus. *Proc. RADAR 2005*, Arlington, Virginia, USA, pp. 747-752.

Nekrasov, A., 2005b. Measuring the sea surface wind vector by the Doppler navigation system of flying apparatus having the track-stabilized four-beam Antenna. *Proc. APMC 2005*, Suzhou, China, Vol. 1, pp. 645-647.

Ouchi, K., 2000. A theory on the distribution function of backscatter radar cross section from ocean waves of individual wavelength. *IEEE Trans. Geosci. Remote Sens.*, 38(2) pp. 811-822.

Sosnovskiy, A.A. and Khaymovich, I.A., 1987. *Radio-Electronic Equipment of Flying Apparatuses*. Transport, Moscow, USSR, 256 p., in Russian.

Wismann, V., 1989. Messung der Windgeschwindigkeit über dem Meer mit einem flugzeuggetragenen 5.3 GHz Scatterometer, Dissertation zur Erlangung des Grades eines Doktors der Naturwissenschaften. Universität Bremen, Bremen, Germany, 119 S.

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to Prof. Dr. Carlo Ferrari, Prof. Dr. Renata Archetti, the Interdepartmental Center of Research in Environmental Sciences, and the University of Bologna for their research opportunity provided, and to the Triple I Consortium for its a postdoctoral grant in the framework of the European Community Mobility Program Erasmus Mundus External Cooperation Window.

AERIAL PHOTO BUILDING CLASSIFICATION BY STACKING APPEARANCE AND ELEVATION MEASUREMENTS

T.T. Nguyen, S. Kluckner*, H. Bischof, F. Leberl

Institute for Computer Graphics and Vision, Graz University of Technology, Austria – {thuy,kluckner,bischof,leberl}@icg.tugraz.at

KEY WORDS: Vision, Urban, Combination, Classification, Colour, DEM/DTM, Model, High resolution

ABSTRACT:

Remote Sensing is trending towards the use of greater detail of its source data, advancing from ever better resolving satellite imagery via decimeter-type aerial photography towards centimeter-level street-side data. It also is taking advantage of an increase in methodological sophistication, greatly supported by rapid progress of the available computing environment. The location awareness of the Internet furthermore demonstrates that large area remote sensing strives for a model of human scale detail. This paper addresses the task of mapping entire urban areas, where objects to be mapped are naturally three dimensional. Specifically we are introducing a novel approach for the segmentation and classification of buildings from aerial images at the level of pixels. Buildings are complex 3D objects which are usually represented by features of different modalities, i.e. visual information and 3D height data. The idea is to treat them in separated processes for learning and then integrate them into a unified model. This aims to exploit the discriminative power of each feature modality and to leverage the performance by fusing the classification potentials at a higher level of the trained model. First, representative features of visual information and height field data are extracted for training discriminative classifiers. We exploit powerful covariance descriptors due to the low-dimensional region representation and the capability to integrate vector-valued cues such as color or texture. Then, a stacked graphical model is constructed for each feature type based on the feature attributes and classifier's outputs. This allows to learn inter-dependencies of modalities and to integrate spatial knowledge efficiently. Finally, the classification confidences from the models are fused together to infer the object class. The proposed system provides a simple, yet efficient way to incorporate visual information and 3D data in a unified model to learn a complex object class. Learning and inference are effective and general that can be applied for many learning tasks and input sources. Experiments have been conducted extensively on real aerial images. Moreover, due to our general formulation the proposed approach also works with satellite images or aligned LIDAR data. An experimental evaluation shows an improvement of our proposed model over several traditional state-of-the-art approaches.

1. INTRODUCTION

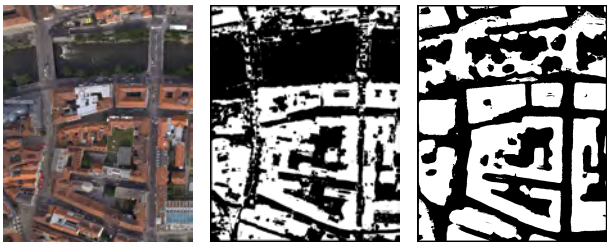
Remote sensing is rapidly moving towards half-meter satellite imagery, decimeter aerial imagery and centimeter-type street-side photography, and all of these in a multi-spectral mode. Simultaneously, large areas of the World are now being mapped at human-scale detail to support the Internet's recent appetite for location awareness. This is resulting in a new domain of large-area urban mapping, however to result not in photo-textured point clouds, but in interpreted objects from which one can build a model of the urban World. A central task is the segmentation and classification of images of buildings. This needs to be fully automated to be at sufficiently low cost so that large area mapping is feasible.

A large urban area may encompass 150 to 500 square-kilometers. Large scale aerial imagery may be at a pixel size of 10 cm. Such large urban area may be covered by 10,000 large-format aerial photographs at high overlaps. We are thus addressing a challenging task of scene interpretation and understanding. It is essential for many location-based applications, such as detailed image description (Meixner and Leberl, 2010), realistic 3D building modeling (Zebedin et al., 2008) or virtual city construction (Leberl et al., 2009). Over the years, the automated building extraction has been an active research topic. Considering a large scale processing, the problem of building classification becomes very difficult for many reasons. Buildings are complex objects with many architectural details and shape variations. Buildings are located in urban scenes that contain various objects from man-made to natural ones. Many of those are in close proximity or disturbing, such as parking lots, vehicle, street lamps, trees, etc. Some objects are covered with shadows or cluttered. These difficulties make the problem of a general building detection challenging. Figure 1 depicts typical urban scenes taken from three challenging datasets *Graz*, *Dallas* and *San Francisco* showing some of these difficulties. We therefore propose an approach which combines several feature cues such as color, texture and 3D information in order to obtain a reliable building extraction from aerial images.

Figure 1: Typical color images of complex urban scenes taken from the dataset *Graz*, *Dallas* and *San Francisco*.

With the success of the aerial imaging technology, high resolution images can be obtained cost-effectively. Multiple sources of data become available, i.e. color, infrared and panchromatic images (Zebedin et al., 2006). Furthermore, since the aerial images are taken with a high overlap from different camera viewpoints, a dense match approach (Klaus et al., 2006) can be applied to obtain range images, representing digital surface models (DSM), from neighboring images. Taking into account the DSM, 3D height information describing the real elevation of each pixel from ground can be computed. The obtained 3D information in combination with visual cues can be exploited efficiently for tasks like accurate building extraction. Figure 2 shows two classification results obtained for a scene of *Graz* by separately incorporating color and 3D information. It is obvious, that a combination of both cues will provide an improved classification results. Moreover, aerial images contain a huge amount of data, which requires efficient methods for processing. This work presents a general

* Corresponding author.



(a) (b) (c)

Figure 2: A classification result obtained by using only color (b) or 3D height information (c). In this work we apply a stacked graphical model in order to combine the strength and advantages of both pipelines.

yet efficient approach that integrates the power of discriminative feature cues.

Automatic building classification and extraction has been a very active research topic in photogrammetry and computer vision for years. The proposed approaches heavily differ in the use of data sources, extracted feature types, the applied models or the evaluation methods (Jaynes et al., 2003, Matei et al., 2008, Lafarge et al., 2008, Mueller and Zaum, 2005, Persson et al., 2005, Xie et al., 2006, Sirmacek and Unsalan, 2008).
Traditional approaches for general image classification problems are mainly based on locally extracted features and a learned classifier that discriminate the object from background. Visual information describing the appearance, such as color and texture are mixed together in a single feature vector to represent the object instance. A concatenating of multiple feature types into a single vector may cause an encountered affect, i.e. one feature type may inhibit the performance of another; besides, it may also cause the problem of over-fitting due to redundancy and correlation in the input data (Duda et al., 2001). Moreover, standard learning algorithms, such as Naive Bayes, logistic regression, support vector machines (SVM) assume that the training data is independent and identically distributed. This is inappropriate in many cases, as image pixels possess dependencies, e.g. if a pixel is labeled as building, it is likely that a neighboring pixel is also labeled as building; non-building pixels tend to be next to other non-building pixels. The spatial dependencies should be exploited properly to improve the classification performance rather than classifying each of the image sites independently.
There have been wide research interests in random field models, i.e. Markov random field (MRF), conditional random field (CRF) (Lafferty et al., 2001, Li, 2001), and their variants in the computer vision community. These models aim to incorporate contextual information to the decision of the object class for improving the performance of the classifiers. In (Korc and Foerstner, 2008), the authors employed MRFs and showed that parameter learning methods can be improved and that using the approach to interpret terrestrial images of urban scenes is feasible. In the vision community, modern approaches exploit graphical models for integrating additional information about the content of a whole scene (Shotton et al., 2006, Larlus and Jurie, 2008, Verbeek and Triggs, 2007).
Recently, Ma and Grimson (Ma and Grimson, 2008) proposed a coupled CRF model for decomposing the feature space in order to learn the object classes. Besides, there have been attempts to model contextual interactions by employing related predictions in a stacked graphical model (SGM) learning framework (Kou and Cohen, 2007). This model enables efficient learning and inference. Moreover, the concept of a relational template can be flexibly exploited to incorporate multi-modal interactions. Our work can be considered as an extension of both, the coupled CRF model (Ma and Grimson, 2008) and the SGM learning (Kou and Cohen, 2007). In this work we propose a novel approach based on an ensemble of SGM in order to integrate different data sources for building classification at the pixel level.
In contrast to the work of Matikainen et al. (Matikainen et al., 2007), where they proposed to use a DSM segmentation and a classification pipeline discriminating buildings from trees, we focus on a more direct and general approach. Our model is comprised of multiple classifiers that are learned over stages and then fused together. Each classifier is responsible for a certain feature modality and modeled as a SGM for the learning procedure. For each SGM, we propose to use a relational template which takes into account the predictions not only of related instances of a certain feature type, but also predictions from other types. This enables to learn not only spatial knowledge of object class, but also the inter-modality dependencies. The proposed system provides a simple yet efficient framework to model a complex object class such as buildings and exploit the potentials from different aspects of the object properties. Learning and inference are effective, general and straightforward, that can be easily applied for many other learning tasks.

Our paper is organized as follows: In Section 2. we introduce our novel framework. Section 3. describes the aerial imagery and the involved feature cues. Section 4. highlights the experimental evaluation. Finally, Section 5. concludes the work and discusses open issues for future work.

2. OUR FRAMEWORK

Let the observed data from an input image be $X = \{\mathbf{x}_i, 0 < i < |X|\}$, where $\mathbf{x}_i$ is the data from a site i. The problem is to find the most likely configuration of the labels $Y = \{\mathbf{y}_i\}$, where $\mathbf{y}_i \in \{c_1 \dots c_k\}$. For an image labeling, a site is a pixel location, and a class may be a car, a building, etc. For the task of the building segmentation each pixel in the aerial image, represented by a feature vector $\mathbf{x}_i$, is mapped to a bit $\mathbf{y}_i \in \{-1, +1\}$, corresponding to either building or non-building. A traditional CRF with local potentials and pairwise (spatial) dependencies can be written as

$$P(Y|X) = \frac{1}{Z(X)} \prod_{i \in S} A(\mathbf{y}_i, X) \prod_{i,j \in N_i} B(\mathbf{y}_i, \mathbf{y}_j, X), \qquad (1)$$

where $A(\mathbf{y}_i, X)$ corresponds to the local potential of $\mathbf{x}_i$ given a class label $\mathbf{y}_i$; $B(\mathbf{y}_i, \mathbf{y}_j, X)$ is the interaction potential function which encodes the dependencies between data X, labels at i and its neighbor j, based on the set of pixels in a neighbor N_i of $\mathbf{x}_i$. $Z(X)$ is a partition function and S defines a set all available image sites. Note that in the formal CRF formulations, potentials depend on the whole image X, not only on the local site $\mathbf{x}_i$. A SGM can be seen as a simplified form of the CRF, given in Equation 1, which allows a flexible structure for the interactions and provides efficient learning and inference. The general stacked model is formulated as a combination of T multiple components of conditional distribution that capture contextual information

$$P(Y|X) = \frac{1}{Z(X)} \prod_{t=1}^{T} p^t(\mathbf{y}_i|X). \qquad (2)$$

The number of components T depends on the model built for a particular application. The flexibility of a CRF formulation al-

lows to incorporate multiple aspects of data from the image, such as: local statistic of an image site, neighboring labels, or potentials from higher levels of contexts. This property will be employed in our framework, where we propose to use a two-staged approach.

2.1 The Ensemble Model - Stage 1

We decompose the input feature space into different feature cues, which represent multiple modalities of the input data. These types of features may be representative for color, texture and 3D information. Assuming that T feature types are extracted from an input image X, $X = \{X^t\}$, $t \in T$, then the CRF of the first stage can be modeled as combination of multiple sources

$$P_{(1)}(Y|X) = \frac{1}{Z(X)} \prod_{t=1}^{T} p^t_{(1)}(Y|X^t), \tag{3}$$

where each $p^t_{(1)}(\mathbf{y}|\mathbf{x})$ is a SGM. The main reasons of decomposing the input data and using an ensemble model are: First, as it has been investigated, object properties such as color, shape, texture, 3D data, etc. play different roles in distinguishing object classes (Kluckner et al., 2009, Ma and Grimson, 2008). Therefore, we treat them separately in different classification processes and combine them at later stages to infer the object classes (Nilsback and Caputo, 2004). At the first stage, we employ multiple strong classifiers learned from different feature modalities. The classifiers provide a probabilistic class assignment in terms of a likelihood. In this work, we use efficient randomized forests (RF) (Breiman, 2001) as base classifiers to generate initial yet accurate confidence maps (Kluckner et al., 2009, Shotton et al., 2008). However, any other types of classifiers, that results a class probability, e.g. boosting, SVM, etc., can be applied to our framework. In order to train the the classifiers of the ensemble, we use fast covariance matrix descriptors as feature representation as proposed by (Kluckner et al., 2009). The details for the feature representation and the base classifier are described in Section 2.3 and Section 2.4.

Since random field modeling approaches exploit contextual information to improve the detection rate of standard classifiers, it is intuitively sensible that different object's property have their own context where it is more likely to appear. This is especially true in our application, where multiple sources of aerial image data are employed, i.e. color image and height data: it may be claimed that pixels with similar color could have similar labels; however, this is not true for height data: pixels with the same height values may belong to buildings or trees. So, the ensemble model comprised of multiple SGMs, where each responses to a certain feature type, is useful to exploit potential of each feature type and its own context.

2.2 The Ensemble of a SGM - Stage 2

We are interested in a model that captures the dependencies among different kinds of feature modalities and spatial knowledge as contexts. Therefore, the second stage of the model is based on the features and the outputs of classifiers from the first stage. Again, we treat each type of features separately. At this stage, we model the dependencies between the feature types and the spatial context. This enables to handle inter-features dependencies and to learn the interactions at a higher level:

$$P_{(2)}(Y|X) = \frac{1}{Z(X)} \prod_{t=1}^{T} p^t_{(2)}(Y|X^t, p_{(1)}). \tag{4}$$

We propose a new relational template for the SGM, in which each feature vector of a certain type is expanded with predictions from its related instances. In particular, each original feature vector of a certain feature type is augmented (stacked) with the prediction confidences from its neighboring sites and confidences from other feature types, which forms a new training set. This allows to learn the spatial dependencies as well as the inter-modality relationships. We use an aggregate function to build the new training dataset: For each feature type $t \in T$, each instance $\mathbf{x}^t_i$ is expanded with the prediction confidences from its neighbor N_j and from other feature types $p^t_{N_j}$ and $p^{T\setminus t}_j$, respectively:

$$\mathbf{x}^t_{i,new} = (\mathbf{x}^t_i, p^t_{N_j}, p^{T\setminus t}_{N_j}, 0 < j < 8). \tag{5}$$

Multiple discriminative probabilistic classifiers are now learned on these new training sets. At this stage, we use a linear SVM due to efficiency and its discriminative power. Finally, the classification confidences of the classifiers are fused together for the inference of the object classes.

2.3 Feature Representation

In the vision community, covariance matrix based descriptors are widely used for detection and classification tasks (Tuzel et al., 2006) due to providing a compact and low-dimensional feature representation. A set of independent feature vectors $\mathbf{f}_i \in F$, where F is an image structure that includes the feature attributes for e.g. color, height, etc. and i defines an image site, can be represented by a sample mean μ_i and a covariance matrix Σ_i defining the first and second order statistics. Importantly, extended integral images (Tuzel et al., 2006) provide an efficient computation of covariance matrices within rectangular image regions. Since the space of covariance matrices is non-Euclidean, these descriptors can not be directly used as a feature representation for learning RFs or SVMs. Here, we exploit a derived representation based on Sigma Points (Kluckner et al., 2009) to obtain a valid feature space, that can be trained with our machine learning techniques. Please note, due to our general model, any other feature representation or classification procedure can be applied.

2.4 Randomized Forests as Base Classifier

An RF classifier (Breiman, 2001) consists of a collection of K decision trees. The nodes of each tree include fast binary decisions that give the direction of splitting left and right down the tree until a leaf node is reached. Each leaf node $l \in L$ contains a learned class distribution $P(c|L)$. By propagating single class distributions bottom-up to the root node for all K trees in a forest the resulting accumulated probabilities yields an accurate class distribution $P(c|L) = \frac{1}{K}\sum_{i=1}^{K} P(c|l_i)$. As demonstrated in (Shotton et al., 2008, Kluckner et al., 2009), RF classifiers give robust and accurate results in challenging image classification tasks. To grow each tree of the forest, node tests are learned by using only a small chosen subset of the training data X^t (Shotton et al., 2008). The learning proceeds from the root node top-down by splitting the subset at each node into tiled left and right subsets. The decisions in the nodes minimize the sample weighted information gain ratio (Shotton et al., 2008) of the class distribution in currently available subsets of the training data. Proposed decisions

in (Shotton et al., 2008) arise by reason of comparing two randomly chosen elements of a given feature vector. At evaluation time, the class distribution for each pixel is determined by parsing down the extracted feature representation in the forest. RFs provide robust probabilistic outputs and are extremely fast to train and test.

2.5 Learning and Inference

In this work, the RF classifiers provide the local class potentials for each individual feature type. After obtaining initial classification confidences at the first stage, the new training sets are constructed using Equation 5. At the second stage, linear SVMs are employed to train these new expanded datasets. We keep the same parameters for individual classifiers at each stage. Thus, there is no need for a parameter tuning in a high-dimensional space. After learning, the classifiers are applied to various test images. However, due to spectral differences in color and specified height conditions, we have to train individual models for each dataset of *Graz*, *Dallas* and *San Francisco*. We then combine the confidence maps (rather than hard output of classifiers) to infer the final object class. Our ensemble model enables to classify buildings from aerial images and to segment building's regions at the pixel level. This involves inferring a true label for each pixel, which is done by computing the most likelihood $y^* = argmax_y P(Y|X)$, given the features X and the classification function. The overall procedure for learning and inference of the model is summarized in Algorithm 1.

Algorithm 1 Learning and Inference

Learning algorithm:
Given a training set (X^t, Y), for each feature types $t \in T$.
For each feature type $t \in T$
- Stage 1: Learn the local model using an RF with (X^t, Y)
- Compute a probabilistic class assignment $p^t_{(1)}$
- Expand the dataset by stacking (Eq. 5): $x^t_{i,new} = (\mathbf{x}^t_i, p_{(1)})$
- Stage 2: Learn the SGM using SVM with (X^t_{new}, Y)

Inference:
Given a test image X, for each feature type $t \in T$
Compute $P^t_{(1)}(Y|X)$ and $P^t_{(2)}(Y|X)$
Infer final class labels: $y^* = argmax_y \prod_t P^t_{(2)}(Y|X)$

3. AERIAL IMAGERY

We perform experiments on high resolution aerial images extracted from three datasets (*Graz*, *San Francisco* and *Dallas*) showing different characteristics. The dataset *Graz* shows a colorful appearance with challenging buildings, the images of *San Francisco* have suburban occurrence in a hilly terrain and *Dallas* includes large building structures and is mainly dominated by gray valued areas. The aerial images are taken with the *Microsoft UltraCam* in highly overlapping strips, where each image has a resolution of 11500×7500 pixels with a ground sampling distance of approximately 10 cm. We use two types of image information, which are: the RGB color image and the 3D height data produced by using the DSM (Klaus et al., 2006) and a subsequently computed digital terrain model (DTM) (Champion and Boldo, 2006). By combining the DTM and DSM, we obtain an absolute elevation per pixel from ground, which is used as the 3D height information. Additionally, we exploit texture information, provided by processing the color images with first-order derivative filters. Figure 3 shows a typical scene taken from *Graz*, including the color image, the hand-labeled ground truth mask and the corresponding normalized 3D information. In our approach we exploit such ground truth map with two classes to train our classifiers in a supervised manner.

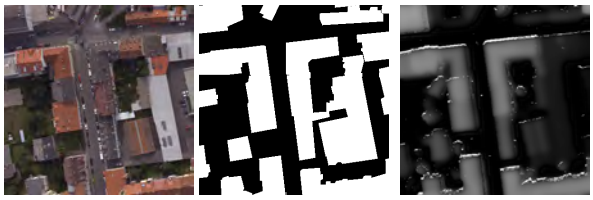


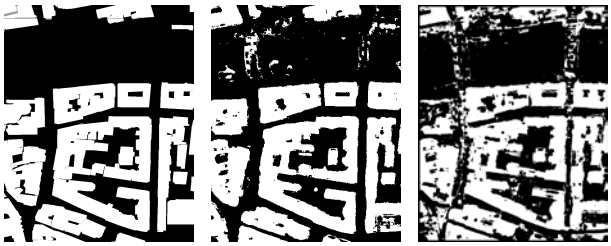
Figure 3: A scene taken from the *Graz* dataset: the color image, the hand-labeled ground truth map and the corresponding normalized 3D height information (from left to right).

Graz	Classification Accuracy (%)		
Model types	Overall	Building	Non-Build.
SVM	88.15	91.47	85.77
RF	85.42	76.95	91.46
Stacked RF model	88.39	91.45	88.39
Our SGM model	**91.65**	93.38	91.09
Dallas	Classification Accuracy (%)		
Model types	Overall	Building	Non-Build.
SVM	93.11	90.40	94.41
RF	91.76	75.86	99.39
Stacked RF model	93.31	90.94	94.63
Our SGM model	**93.76**	90.81	95.12
San Francisco	Classification Accuracy (%)		
Model types	Overall	Building	Non-Build.
SVM	87.97	81.31	96.79
RF	91.17	89.33	93.62
Stacked RF model	92.12	88.34	94.32
Our SGM model	**93.98**	94.40	93.42

Table 4: Performance evaluation of different models in terms of correctly classified pixels obtained for the datasets *Graz*, *Dallas* and *San Francisco*. We compute a global rate and the accuracy individually for each of the classes building and non-building by considering a hand-labeled ground truth map. The accuracy measurements for the building and non-building class are also referred to as completeness and correctness, respectively. The rates are given for the models integrating the visual feature cues and the 3D height information.

4. EXPERIMENTS

In this section we evaluate our proposed framework on a large amount of real world data. We compare the performance of our model to several traditional state-of-the-art approaches. The comparisons include the performances of a traditional RF and SVM classifiers, both integrating appearance and 3D height, a SGM with RFs as base classifier (in the following we call it a stacked RF), and our ensemble model also including the second stage of our approach. Each of the base RF classifiers consists of $K = 8$ trees with a maximum depth of 14. For the stacked models (including the stacked RF and our ensemble model), the cross-validation parameter is set to 4 and the relational template takes into account 8 direct neighboring pixel sites. We use a linear SVM for learning the stacked RF and our ensemble model at the second stage. The feature instances are collected on a regular image grid incorporating a small spatial neighborhood of 11 pixels in order to include important context information. The covariance feature representation based on Sigma Points comprises a compact statistical description of an image region with a dimension of $d' = d(2d + 1)$, where d denotes the number of considered feature modalities. Note, we consider each color channel of an RGB image as a single modality.



(a) (b) (c)

Figure 5: The corresponding classification results for the images presented in Figure 2 by integrating appearance and height information: (a) the ground truth image, (b) the results obtained by our SGM and (c) a traditional RF based classification. Compared to Figure 2, our SGM obtains sharp delineated building boundaries and false positive classified regions are consistently eliminated.

For training and testing the model, six representative triplets extracted from the large images are used. Each of these sub-images has a size of 4000×3200 pixels. The images cover large dense urban areas, which contain various complex objects, such as buildings of variant sizes and complex architectures, road net, parking lots, trees, shadow, water surface, etc. For the quantitative comparison we evaluate each labeled pixel with respect to the available ground truth data.

Considering Figure 2, it is obvious that a classifier, only trained on 3D features, fails in river regions, where the dense matching process provides regions with many undefined heights. In addition, tree areas are classified as buildings due to similar building height. By using only the 3D height data an RF classifier obtains a detection rate of 78.57% on this scene extracted from *Graz*. Exploiting only the visual feature modality, the raw RF classifier correctly assigns the pixels at a rate of 79.20%. However, there are significant missed detections in regions on the ground, that have similar appearance as buildings. A combination of the height and the visual features within an RF classification process significantly improves the final labeling at the pixel level to an accuracy of 85.42%. By integrating the height field data with the visual information within our proposed SGM framework, we obtain an overall pixel classification rate of more than 90% on all three datasets. The detection rates in terms of accuracy at the pixel level of different models are summarized in Table 4. The supervised segmentation of building regions obtained by a traditional RF classifier is shown in Figure 5(c), while the performance of our SGM is depicted in Figure 5(b).

The classification is given as raw outputs of each model without applying a post-processing step. However, this could be done to remove small noisy areas on the ground. Besides, our SGM obtains sharp delineated building boundaries and false positive classified regions are consistently eliminated. The improvement is obvious and results from the feature decomposition and integration at higher level with spatial context. Moreover, we obtain a very fast learning and inference thanks to the intrinsic simple model structure and the efficient relational template for the stacked graphical learning. A classification of an image with a dimension of 4000×3200 pixels can be obtained within few minutes using an unoptimized implementation. Figure 6 shows an improved performance of our approach compared to traditional state-of-the-art methods such as RF classifiers on larger scenes taken from *Dallas* and *San Francisco*, respectively.

5. CONCLUSION

We have proposed an efficient approach for learning multiple feature modalities, i.e. visual features and 3D height data. Our method decomposes an input feature space into different feature modalities in order to train individual probabilistic classifiers. In this work we used randomized forests as base classifiers, trained with various feature types, at the first stage of a stacked graphical model. Then, an ensemble of stacked models with a novel relational template has been employed for learning the dependency of different modalities. We successfully applied the proposed model to the challenging problem of the building classification task in high resolution aerial images, taken from three different datasets. Experiments have shown an improvement of our approach over several traditional state-of-the-art methods. The model is suitable for learning 3D objects like buildings from aerial imagery, but can be applied for other object classes. Due to efficiency, the proposed framework provides a promising application for large-scale computation in aerial imagery. For future work there should be more study on modeling context information for each feature type, which represent different aspects of data. Multiple kernels would be helpful in weighting the contribution of each source of information. In addition, we plan to apply our framework to various detection tasks in standard evaluation image collections.

REFERENCES

Breiman, L., 2001. Random forests. In: Machine Learning, pp. 5–32.

Champion, N. and Boldo, D., 2006. A robust algorithm for estimating digital terrain models from digital surface models in dense urban areas. In: Proceedings International Society for Photogrammetry and Remote Sensing.

Duda, R. O., Hart, P. E. and Stork, D. G., 2001. Pattern Classification. New York: Wiley.

Jaynes, C., Riseman, E. and Hanson, A., 2003. Recognition and reconstruction of buildings from multiple aerial images. Computer Vision and Image Understanding 90(1), pp. 68–98.

Klaus, A., Sormann, M. and Karner, K., 2006. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. Proceedings International Conference on Pattern Recognition.

Kluckner, S., Mauthner, T., Roth, P. M. and Bischof, H., 2009. Semantic classification in aerial imagery by integrating appearance and height information. In: Proceedings Asian Conference on Computer Vision.

Korc, F. and Foerstner, W., 2008. Interpretation terrestrial images of urban scenes using discriminative random fields. In: Proceedings International Society for Photogrammetry and Remote Sensing.

Kou, Z. and Cohen, W. W., 2007. Stacked graphical models for efficient inference in markov random fields. In: Proceedings SIAM International Conference on Data Mining.

Lafarge, F., Descombes, X., Zerubia, J. and Pierrot Deseilligny, M., 2008. Automatic building extraction from dems using an object approach and application to the 3d-city modeling. International Journal of Photogrammetry and Remote Sensing 63(3), pp. 365–381.

Lafferty, J., McCallum, A. and Pereira, F., 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: Proceedings International Conference on Machine Learning.

Larlus, D. and Jurie, F., 2008. Combining appearance models and markov random fields for category level object segmentation. In: Proceedings Computer Vision and Pattern Recognition.

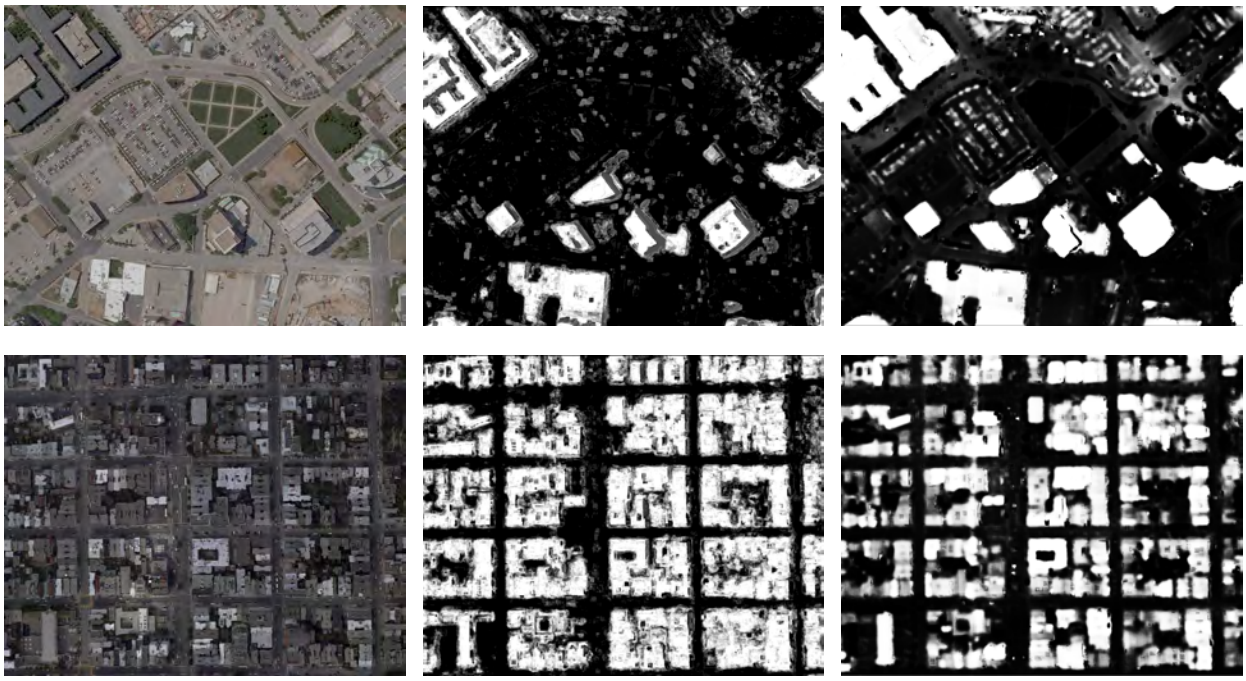


Figure 6: A visual comparison between the classification results of an RF classifier (second column) and our ensemble model (third column). The first row shows the building classification, given in terms of SVM output confidence values, for a scene of *Dallas*, while the second row depicts the results for a part of *San Francisco*.

Leberl, F., Kluckner, S. and Bischof, H., 2009. Collection, processing and augmentation of vr cities. In: Proceedings 52. Photogrammetric Week.

Li, S. Z., 2001. Markov random field modeling in image analysis. Springer-Verlag New York, Inc.

Ma, X. and Grimson, W., 2008. Learning coupled conditional random field for image decomposition with application on object categorization. In: Proceedings Computer Vision and Pattern Recognition.

Matei, B., Sawhney, H., Samarasekera, S., Kim, J. and Kumar, R., 2008. Building segmentation for densely built urban regions using aerial lidar data. In: Proceedings Computer Vision and Pattern Recognition.

Matikainen, L., Kaartinen, K. and Hyyppae, 2007. Classification tree based building detection from laser scanner and aerial image data. In: Proceedings International Society for Photogrammetry and Remote Sensing, Workshop Laser Scanning.

Meixner, P. and Leberl, F., 2010. Describing buildings by 3-dimensional details found in aerial photography. In: Symposium "100 Years ISPRS - Advancing Remote Sensing Science".

Mueller, S. and Zaum, D. W., 2005. Robust building detection in aerial images. In: Proceedings International Society for Photogrammetry and Remote Sensing, Workshop CMRT.

Nilsback, M. and Caputo, B., 2004. Cue integration through discriminative accumulation. In: Proceedings Computer Vision and Pattern Recognition.

Persson, M., Sandvall, M. and Duckett, T., 2005. Automatic building detection from aerial images for mobile robot mapping. In: Proceedings Symposium on Computational Intelligence in Robotics & Automation.

Shotton, J., Johnson, M. and Cipolla, R., 2008. Semantic texton forests for image categorization and segmentation. In: Proceedings Computer Vision and Pattern Recognition.

Shotton, J., Winn, J., Rother, C. and Crimininsi, A., 2006. TextonBoost: Joint Appearance, Shape and Conext Modeling for Muli-class object Recognition and Segmentation. In: Proceedings European Conference on Computer vision.

Sirmacek, B. and Unsalan, C., 2008. Building detection from aerial images using invariant color features and shadow information. In: Proceedings International Symposium on Computer and Information Sciences.

Tuzel, O., Porikli, F. and Meer, P., 2006. Region covariance: A fast descriptor for detection and classification. In: Proceedings European Conference on Computer Vision.

Verbeek, J. and Triggs, B., 2007. Region classification with markov field aspect models. In: Proceedings Conference on Computer Vision Pattern Recognition.

Xie, M., Fu, K. and Wu, Y., 2006. Building recognition and reconstruction from aerial imagery and lidar data. In: International Conference on Radar.

Zebedin, L., Bauer, J., Karner, K. F. and Bischof, H., 2008. Fusion of feature- and area-based information for urban buildings modeling from aerial imagery. In: Proceedings European Conference on Computer Vision.

Zebedin, L., Klaus, A., Gruber-Geymayer, B. and Karner, K., 2006. Towards 3d map generation from digital aerial images. International Journal of Photogrammetry and Remote Sensing 60(6), pp. 413–427.

ACKNOWLEDGEMENTS

This work was financed by the FFG Project APAFA (813397) and the Austrian Science Fund Project W1209 under the doctoral program Confluence of Vision and Graphics.

MULTITEMPORAL RADARSAT-2 POLARIMETRIC SAR DATA FOR URBAN LAND-COVER MAPPING

X. Niu [a], Y. Ban [a, *]

[a] Division of Geoinformatics, Department of Urban Planning and Environment,
Royal Institute of Technology, SE-10044 Stockholm, Sweden- (xin.niu, yifang.ban)@abe.kth.se

KEY WORDS: Land cover, classification, SAR, hierarchical, multitemporal

ABSTRACT:

The objective of this research is to evaluate multi-temporal RADARSAT-2 polarimetric SAR data for urban land-cover classification using a novel classification scheme. Six-date RADARSAT-2 Polarimetric SAR data in both ascending and descending orbits were acquired during June to September 2008 in the rural-urban fringe of the Greater Toronto Area. The major land-cover types are built-up areas, roads, golf courses, forest, water and several types of agricultural crops. In this research, the different urban land-cover types and their corresponding polarimetric behaviors were studied. The polarimetric signatures of the various urban land-cover types were extracted from the RADARSAT-2 SAR images and analyzed using a new hierarchical multitemporal classification method. The results showed that the new classification method yielded high classification accuracy, with overall accuracy of 82.1% and Kappa coefficient 0.80 for the major 11 land-cover classes. The classification scheme can effectively extract the urban structures by mapping urban related classes such as streets and major roads with the higher user's accuracy, which is difficult to achieve using a single-date data.

1. INTRODUCTION

In 2008, the world crossed an invisible but momentous milestone: the point at which more than half the people on the planet living in cities (The World Watch Institute, 2007). Urbanization and the impact of human settlements are two of the main causes of global environmental degradation. Therefore, mapping and monitoring urban landuse/land-cover and their changes in a timely and accurate manner is of critical importance for sustainable urban planning and environment protection (Ban *et al*., 2010). With the launch of advanced remote sensors in recent years such as RADARSAT-2 SAR and TerraSAR-X, multi-temporal, high-resolution, polarimetric SAR data are becoming routinely available for surveying fast expanding urban areas.

Comparing with the single-polarization SAR data, Polarimetric SAR data provide the description of the land features from the observations of various polarizations. Thus, more information can be explored for classification. In the literature, the studies of the polarimetric SAR data focus on the generation of the efficient descriptors from the scattering coefficients and introducing those descriptors to the conventional classification methods. The Cloude/Pottier decomposition (Cloude and Pottier, 1997) and freeman decomposition (Cloude and Pottier, 1996) are well-known examples.

To improve the classification accuracy, multitemporal data have been used whenever the data are available (e.g., Ban & Howarth, 1999; Goodenough & Chen, 2005; Waske *et al*., 2006). The advantages of the multitemporal classification could be given in the following aspects. (1) The rarely changed parts in different dates will confirm each other to increase the credibility. (2) The temporal attributes could be exploited for certain classes such as crops. (3) Data from different orbit, for example ascending and descending, will provide complement views of the observed scenes. Various studies have been conducted using multi-temporal and/or polarimetric SAR data. For examples, Galli *et al.* (2007) investigated a joint segmentation technique on a sequence of multi-temporal single-channel SAR images to improve the classification. Tan *et al*. (2007) assessed SVM classifier for classification of crops using multi-temporal Polarimetric SAR data. Park and Chi (2006) investigated a fuzzy logic fusion of multi-temporal multi-polarization SAR data for landcover classification. Shimabukuro *et al*. (2007) led a case study on the mapping of the deforested area using multi date JERS-1 SAR data with HH polarization. Goodenough & Chen (2005) fused the Polarimetric SAR data sets from winter and summer to map the different forest types. Waske *et al*. (2006) used the decision tree to map urban and rural using the multi-date data. Chen *et al*. (2007) explored the feasibility of residential and rural area classification using multi-temporal SAR images using knowledge-based approach.

Few studies, however, used high-resolution polarimetric SAR data for urban analysis due to data availability. And to the best of our knowledge, there is no research on the classification of the high-resolution RADARSAT-2 polarimetric SAR data in urban areas.

Furthermore, using the high-resolution SAR data also poses new challenges on urban land-cover classification due to the complexity of the urban environment. Thereby, it is necessary to exploit the contextual and spatial information as well as knowledge-based approach to improve classification reuslts (Ban and Hu, 2007; Ban *et al*, 2010).

Thus, in this paper, we propose a new rule-based hierarchical classification method, using multitemporal Polarimetric SAR data to map urban land-cover and extract the fine urban

* Corresponding author

structures. The specific objectives are: 1. to evaluate the capacity of the RADARSAT-2 fine-beam polarimetric SAR for the urban mapping; and 2. to investigate the effectiveness of the proposed multitemporal classification schemes.

2. STUDY AREA AND DATA DESCRIPTION

The study area is located in the north and northwest of the Greater Toronto Area (GTA), Ontario, Canada, where rapid urban expansion is underway. The major land use/land-cover classes in the rural-urban fringe area are high-density area, low-density area, wide roads and airports, path, parks, golf courses, forests, water and several types of crops. Those classes were chosen to characterize the complex area.

The data used for this research consist of six-date RADARSAT-2 fine-beam polarimetric SAR (PolSAR) data, which contain the HH, HV, VH and VV polarizations. The centre frequency at this beam mode is 5.4GHz, i.e., C-band and the spatial resolution is 8 meters. The detailed descriptions of these images are given in Table 1.

Data	Orbit	Incident angle range (degree)
Jun. 11 2008	Ascending	40.179~ 41.594
Jun. 19 2008	Descending	40.215~ 41.619
Jul. 05 2008	Ascending	40.182~ 41.597
Aug. 06 2008	Descending	40.197~ 41.612
Aug. 22 2008	Ascending	40.174~ 41.590
Sep. 15 2008	Ascending	40.173~ 41.588

Table 1. RADARSAT-2 Fine-Beam SAR Imagery

The six-date SAR data are acquired during June to September in 2008, when most of the crops experienced the seasons from flourish to harvest. In contrast, the urban area kept stable in such a short term. Two data groups are naturally formed according to their orbit mode. Thus, the June 11, July 05, August 22, September 15 images make up the ascending group while the June 19, August 06 comprise the descending group.

The different orbit modes introduce different views of the observed objects, which can be used to complement each other. However the angles also pose a challenge to image registration due to the different radar look directions.

In spite of that, all the data are collected in the similar incident angles. Hence the multi temporal data within each group could match each other well, which will benefit the multitemporal classification.

3. METHOD

The proposed hierarchical object-based rule-based classification is illustrated in Figure 2. First, the whole scene was segmented at multi scales using the Pauli parameters derived from the filtered raw Polarimetric SAR data. Secondly, land-cover/land use classes were extracted into different feature layers. Finally we combine those layers in hierarchy according to their credibility levels. The multitemporal feature extraction schemes lies in two aspects: 1.The layer of urban and rural is the classification result on the stacked multi-date data. 2. Any other layer is the fusion result of that specific class from multiple single-date classifications.

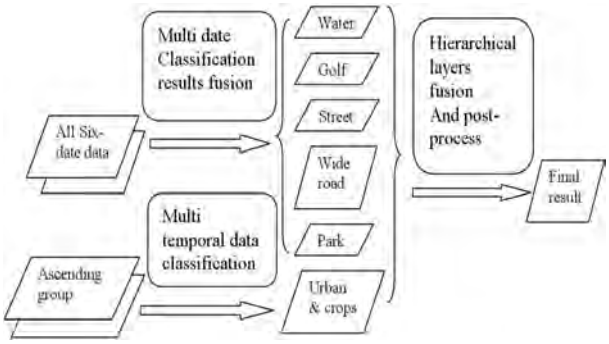


Figure 2. Multitemporal classification scheme

3.1 Orthorectification and Registration

All the six-date RADARSAT-2 fine-beam polarimetric SAR data are orthorectified by the DEM with resolution of 30m. Then all the data are registered to the National Topographic Database (NTDB) vector data. The images could overlay each other well that all the streets are perfectly matched, especially for data in the same orbit mode, i.e., ascending or descending. That excellent overlapping is the base for our multitemporal classification scheme.

3.2 Preprocessing and Pauli decomposition

First, we extract the hermitian coherency matrix <T> for all the raw polarimetric SAR data. The coherency matrix was then filtered by Lee refined filter (Lee, 1981). The number of looks and the window size are set as 2 and 7 respectively. The Pauli parameters are directly obtained as the diagonal elements of the coherency matrix <T>.

The Pauli parameters (Cloude and Pottier, 1996): |HH+VV|, |HH-VV|, |HV| are the measurements of the relative powers of the three physical models: odd-bounce, dihedral oriented at 0 degrees and volume scattering. The real examples for the above models are rough surface, urban building and vegetation respectively. Since the total power of the targets is equal to that of the backscattering matrix. It is often used for illustration.

Although we do not use all the other elements in the coherency matrix <T>, i.e., the real and image parts of (HH+VV)(HH-VV)*, (HH+VV)HV* and (HH-VV)HV* which together represent all the polarization information, the Pauli parameters could present the most contrast between the land-cover/land use classes. In this experiment, the logarithm (base 10) of the Pauli parameters is used as the spectral channels for classification.

3.3 Multi-Scale Segmentation with eCognition

In the experiments, the multiresolution segmentation algorithm in eCogniation is selected as the segmentation method. This segmentation technique aims to locally maximize the homogeneity within the objects. This process, which starts from the pixel level, iteratively aggregates the neighboring candidate segments until they reach the given scale. The homogeneity criterion makes a trade-off between the spectral and spatial domain. In spectral domain, the homogeneity is decided by the

standard deviation, while in the shape domain, there is a balance between the compactness and smoothness. (Baatz *et al*., 2004).

In this research, multi-scale segmentation hierarchy is constructed for each single date data. This hierarchy has three segmentation levels whose scales are 50, 100 and 200. The higher level is the merge result of the lower level. The shape and compact parameters for segmentation are set as 0.4 and 0.5 for the scale 50, 0.5 and 0.5 for the scale 100 and 200. Those empirical segmentation parameters for all the date data are selected by trials and comparisons.

The reasons for using the hierarchical multi-scale segmentation layers are: 1. The land-cover/land use type needs appropriate shape size to be represented. For instance, the water pond, golf course area are directly extracted from the smaller scale/higher level layer, while the streets are from the larger scale/lower level layer. 2. The relationship between the super or sub segmentation layers offers spatial information of the object's surroundings. For example, the mapping of the park benefits from the information from the classification of higher level/ smaller scale segmentation layer.

3.4 Rule-based Class Layer Extraction and Multitemporal Fusion

As mentioned, the class layer of water, street, wide road, golf or park is the fusion of that specific class from multiple single-date classification results. However, the urban and rural areas are classified on the stacked multi-date data. The golf course class is extracted from the segmentation level with scale of 200 and wide road and water from the level of scale 100 while the rest are from the level of scale 50.

The advantages of the hierarchical layers extraction could be in two folds: 1.We can focus on some specific land-cover classes to specifically improve their classification accuracy. 2. Different layer fusion schemes can be employed for different purposes (e.g., to achieve high user's accuracy or producer's accuracy).

The rules involved in the classification consist of criterions considering the objects' data value and shape. On one hand, for the classes such as water, forest, crops and urban which are distinguished by the data value, the standard nearest neighbor classifier is applied to obtain the initial classification results based on the samples. Since the segmentation results are various for different date data, the samples are not exactly the same for all the date data, but they all collected from the same area. On the other hand, the road and street are extracted by the shape characters and the relationships with the neighboring classes. Following are the descriptions of such rules and fusion schemes for each class.

3.4.1 Water: The water class is extracted by standard nearest neighbor classifier with the mean values and the standard deviation on scale 100. On this level, most of the ponds are segmented as an entire part. Thus, the other small low scattering objects like the shadow will be excluded from this level and not mistaken as water.

The water area extracted from the single date has higher accuracy. However, some turbid ponds has the scattered values behaved differently in one looking direction and another. For instance, in figure 3, the pond from June 11 (ascending) has higher scattered values while in June 19 (descending) it has normally low values. Therefore, first we make the intersection of the water layers within each ascending or descending group. Then, the two group intersection layers are combined as the final water layer.

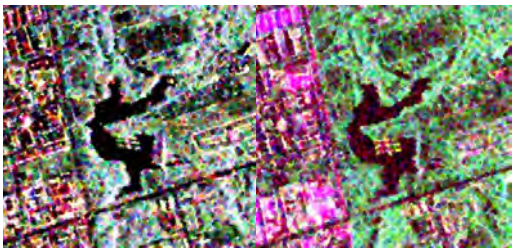


Figure 3. Comparison of the Pauli image of the water pond . From left to right are ponds from June 19 (descending) and June 11(ascending)

3.4.2 Golf Course: The golf course is a complex area consists of grass, pond, bare field and some houses. However, at the higher segmentation scale, those adjacent ingredients can form a distinct part from the surrounding. It can be found from figure 4 that at scale 200, the golf area is well segmented and distinct from the other part. Thus the golf is extracted by standard nearest neighbor classifier with the mean values and the standard deviation on scale 200.

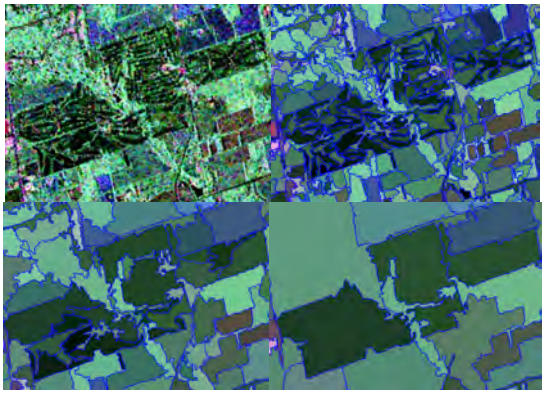


Figure 4. Comparison of segmentation results of different scale size of the golf area. From top left to bottom right are the golf Pauli image from June 19 and its segmentation result of scale 50,100 and 200.

Since the segmentation result is not always perfectly matched the golf area, thus the fusion layers formed by stacking the golf class from all the dates. If in some area the overlap time exceeds certain threshold, this area will be labeled as golf course.

3.4.3 Street: The streets are distinguished on scale 50 by its shape characters and the relationship with the urban area. The street segment should be the narrow winding object which has the higher roundness (defined by the difference of the radius of the smallest enclosing ellipse and the largest enclosed ellipse) and lower rectangular fit (defined as the fitness to the rectangle which has the same area). The width of the skeleton of the street should be smaller. And it should border on more urban objects.

The fusion scheme of the street layer is similar to the golf layer.

3.4.4 Wide road: Similar to the street, the wide road is mainly extracted by the shape characters. Comparing with the street, the wide road has wider width, and less curves. Thus, the longer wider segment will be marked as wide road. Thereby, the rules would consider the ratio of the length to the width, the width range. Furthermore, the flat road usually has lower scattering. Since the wide roads often run through the urban area, it has higher contrast to the brighter neighbor objects.

Those criterions are combined to map the wide road of each single date on the segmentation level of scale 100. Then the results from multi date fuse like the process of the street layer.

3.4.5 Park: The park is defined as the vegetation or bare field embedded in the urban area. Thus the initial classified crops enclosed by the urban patch will be marked as park. However, the scattered odd man made structures in the rural will lead the mistakes by labeling the adjacent crops as park. Thus, we first map those areas with scattered houses embraced by crops as rural on the higher segmentation level of scale 100 to avoid such error. Then the park will be marked at the lower level of scale 50 based on the relationship both to the super and neighboring objects. The multi date result fusion is the same to road.

3.4.6 Rural and urban: The rural and urban area (including the forest) is simply segmented and mapped on the stacked ascending data group, using the standard nearest neighbor classification with the mean value and standard deviation of the data. We do not choose the descending group to distinguish the urban and rural because of its poor performance to discriminate the low density area from the forest. Since the ascending group has four date data from different seasons, the multi-temporal character of crops can be explored to improve the classification accuracy. Figure 5 gives the illustration of 2 kinds of crops in the time sequence of ascending group.

3.5 Hierarchical Fusion Scheme and Post-processing

All the class layers are one by one input as the thematic maps given the rural and urban layer as the base layer to be fused on. While one class layer are being fused, on the base layer, the objects which were covered by the class feature beyond certain portion will be labeled as that class. It is to avoid the scraps as if we directly superimpose the class layers on the base layer, since our class layers are the fusion from different segmentation results. The labeling process is in the order of layer of park, wide road, street, golf and water. The higher the hierarchical layers, the more accurate they are, and less confused with the others. The credibility of the class layers depends on the class' character and the extracting schemes. For example, the park is often confused with street, wide road, thus, it will be mapped below those two layers. The golf course often contains ponds, thus the water layer is on it.

Post-processing consists of two approaches: 1. On the fused map, we strict certain criterions to mark the potential objects for specific classes as we get the fused map with higher accurate context information than the single date data. For example, some potential streets in the fused map may be found and connected. 2. The map is filtered to remove the tiny isolated objects within another class. For example, the shadows or the trunks in the forest area could be removed.

3.6 Accuracy Assessment

For the accuracy test, the Quick-bird images, NTDB vectors and maps are referred. Test areas which contain more than 2000 pixels for each class are randomly selected. The quality of the classification results is assessed by various parameters such as overall accuracy and Kappa coefficient of agreement (or Kappa). They were analyzed to compare classification results with the reference data in confusion matrices.

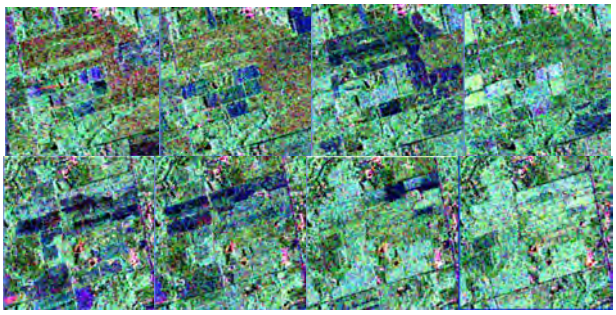


Figure 5. The temporal character of crop 2 (top 4 Pauli images) and crop 3 (bottom 4 Pauli images), from left to right are from June 19, July 05, August 22 and September 15. All from ascending groups.

4. RESULT AND DISCUSSION

For 11 urban land-cover classes, the overall classification accuracy of 82.1% and Kappa coefficient 0.80 are achieved using our method. The confusion matrix (Table 6) shows that HD, LD, Forest, Water and Agriculture are classified very well. Water (96.66%) and low density area (88.42%) achieve very good accuracies. Park (72.08%) and golf course (75.4%) have relatively lower accuracies as they are difficult to be discriminated from crops or water ponds due to the similar backscattering, As a result of our strict fusion rules, however, they could achieve higher user accuracies. The wide roads (80.76%) and streets (80.63%) are well classified to illustrate the urban structures. Figure 7 shows the example of the wide road and street layers.

The character of our multitemporal classification scheme can be seen from the table that the classes of man-made categories have higher user accuracies. Because rules for those class layers are carefully defined in our fusion schemes, like park and golf. Although the wide road has relatively lower user's accuracy (79.54%), it is caused by the vague definition between street and itself. That is why wide road had a large omission error to street. But if we uniformly treat those two classes as road, the accuracy will definitely become much higher.

The park class has the most complicated situation. Not all the vegetation within the enclosed in the urban area are parks, and there are some parks on the fringe of the urban as well. That makes the rules for identifying parks difficult. The crops are mainly confused among themselves and with the park class.

	Reference data (Percent)											
Class	**Water**	**HD**	**LD**	**Crop1**	**Crop2**	**Crop3**	**Road**	**Street**	**Golf**	**Forest**	**Park**	**User's Accuracy**
Water	96.66	0.00	0.00	0.00	0.00	0.00	0.00	1.56	2.12	0.00	0.00	93.11
HD	0.00	82.44	4.79	0.00	0.00	0.00	0.00	1.17	0.00	4.73	0.00	84.76
LD	0.00	14.42	88.42	0.00	0.00	0.00	1.57	9.24	0.00	0.00	0.00	74.97
Crop1	0.00	0.00	0.00	81.27	10.68	20.92	3.97	0.53	15.12	6.35	20.48	31.14
Crop2	0.00	0.00	0.00	8.94	87.48	0.00	0.00	0.00	0.00	0.00	0.00	95.07
Crop3	2.41	0.00	0.00	0.00	1.84	79.08	2.66	0.57	1.74	1.62	0.00	88.62
Road	0.92	1.64	0.00	0.00	0.00	0.00	80.76	0.76	2.57	0.52	5.16	79.54
Street	0.00	0.13	0.00	0.00	0.00	0.00	10.95	80.63	0.00	0.00	2.28	91.81
Golf	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	75.04	0.00	0.00	100.00
Forest	0.00	0.00	0.58	0.00	0.00	0.00	0.00	0.00	2.10	86.78	0.00	97.33
Park	0.00	1.37	6.21	9.79	0.00	0.00	0.00	0.09	5.53	1.31	72.08	80.33
Producer's Accuracy	96.66	82.44	88.42	81.27	87.48	79.08	80.76	80.63	75.04	86.78	72.08	

Table 6. Confusion Matrix.

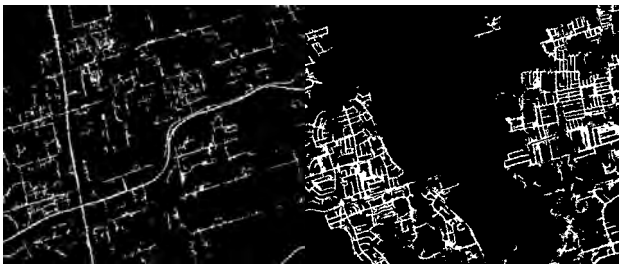


Figure 7. The wide road structure (left) and part of the street layer (right)

Figure 8 illustrates four examples from our classification result compared with the Quickbird images. It is observed that the classification of built-up areas, wide road, street networks, golf course, and agricultural fields match the features in the Quickbird images very well.

Comparing with the previous multitemporal classification methods such as Chen *et al.* (2007) and Park and Chi (2006), the novelty of our method could be summarized in the following way: 1.Our method combines the hierarchical fusion schemes with the multitemporal conception thus could offer robust and high accuracy for specific classes, while the previous method could only treat all the classes with the same criterions. 2. Comparing the other urban studies such as Waske *et al.* (2006), we could offer more classes to reflect the finer urban structures, while most of the previous studies only focus on mapping the whole urban area. And even there are considerations of the structure mapping, their methods are mostly directed for the wide, major roads extraction, while our scheme propose a efficient way enable to explore the finer structures like streets and parks.

5. CONCLUSIONS

RADARSAT-2 fine-beam polarimetric SAR data were evaluated for land-cover mapping in the rural-urban fringe of the Greater Toronto Area. The multi-temporal hierarchical fusion method applied on the six-date data generates higher classification result, with overall accuracy of 82.1% and Kappa coefficient 0.80 for the major 11 land-cover classes including high-density built-up areas, low-density built-up areas, wide roads, street, forests, parks, golf courses, water and several types of agricultural lands. The results indicated that the fusion scheme can effectively extract the urban structure by mapping urban related classes such as streets and major roads with the higher user's accuracy, which is difficult to achieve using a single-date data.

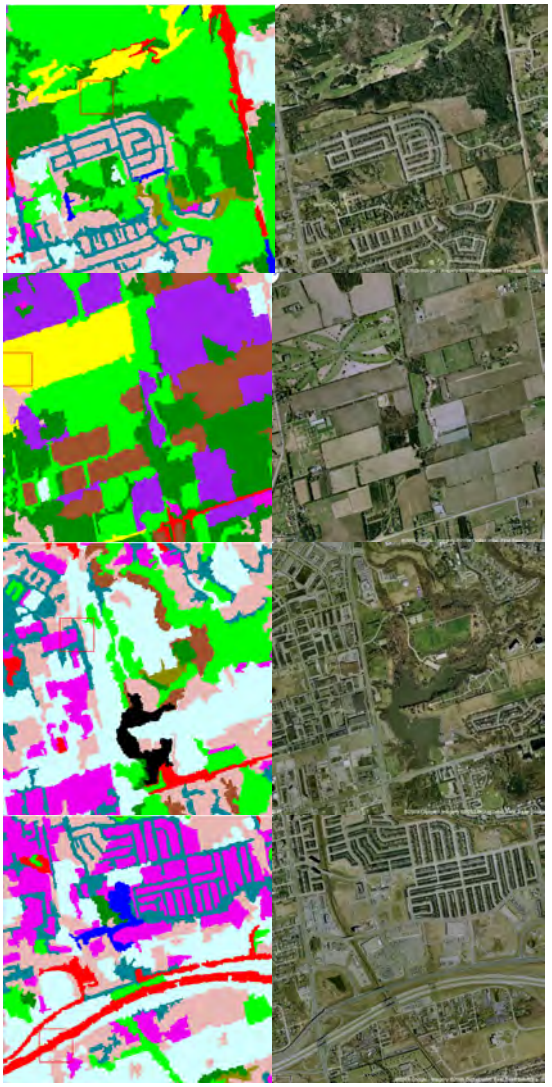


Figure 8. Land-cover classification: selected examples, each comparison consists pair of classification result and the Quick-bird reference map.

REFERENCE

Baatz M., Bentz U., Dehghani S., Heynen M., 2004. eCognition User Guide 4. Definiens Imaging, Muenchen.

Ban, Y. and P.J. Howarth. 1999. Multitemporal ERS-1 SAR data for crop classification: a sequential-masking approach. *Canadian Journal of Remote Sensing*, Vol. 25, No. 5, pp. 438-447.

Ban, Y. and H. Hu, 2007. RADARSAT Fine-Beam Data for Land-Cover Mapping and Change Detection in the Rural-Urban Fringe of the Greater Toronto Area. In: *Urban Remote Sensing Joint Event, 2007*, Paris, France, pp.1-7, 11-13.

Ban, Y., H. Hu and I. Rangel. 2010. Fusion of Quickbird MS and RADARSAT SAR data for urban land-cover mapping: object-based and knowledge-based approach. *International Journal of Remote Sensing*, Volume 31 Issue 6, 1391

Chen, F.L. Wang, C. Zhang, H., 2007. SAR images classification using case-based reasoning method. In: *IEEE International Geoscience and Remote Sensing Symposium,2007*, Barcelona, Spain, pp.2048-2051.

Cloude, S.R. and Pottier, E., 1997. Application of theH/A/alpha polarimetric decomposition theorem for land classification. Proc. SPIE, Vol. 3120, pp.132-143.

Cloude, S.R. and Pottier, E., 1996. A review of target decomposition theorems in radar polarimetry. IEEE Transactions on Geoscience and Remote Sensing (GRS), vol. 34, no. 2, pp. 498–518.

Galli, L., Passaro, D. and Avolio S., 2007. A Multi-scale Joint Segmentation Technique for Multitemporal and Multisource Remote Sensing Images. *Image and Signal Processing for Remote Sensing*, XIII., Volume 6748, pp. 674804.

Goodenough, D.G. and Chen, H., 2005. Multitemporal Evaluation with ASAR of Boreal Forests. In: *IEEE International Geoscience and Remote Sensing Symposium, 2005*, Seoul, South Korea,Vol.3,pp.1662-1665

Lee, J. S., 1981. Refined filtering of images using local statistics. *Computer Graphics and Image Processing*, Vol. 15, pp. 380-389.

Park, N.-W. Chi, K.-H., 2006. Land-cover classification using multi-temporal /polarization C-band SAR data. In: *IEEE International Geoscience and Remote Sensing Symposium, 2006*, Denver, USA, pp.188-191

Shimabukuro, Yosio E., Raimundo Almeida-Filho, Tatiana M. Kuplich and Ramon M. de Freitas, 2007. Mapping and monitoring land-cover in Corumbiara area, Brazilian Amazônia, using JERS-1 SAR multi-temporal data. In: *IEEE International Geoscience and Remote Sensing Symposium, 2007*, Barcelona, Spain, pp. 3370-3373.

Tan, C.-P., J.-Y. Koay, K.-S. Lim, H.-T. Ewe and H.-T. Chuah, 2007. Classification of multi-temporal sar images for rice crops using combined entropy decomposition and support vector machine technique. *Progress In Electromagnetics Research*, PIER 71, 19-39.

The World Watch Instutite, 2007, *State of the World 2007: Our Urban Future* (W. W. Norton).

Waske, B., Schiefer, S. and Braun, M., 2006. Random feature selection for decision tree classification of multi-temporal SAR data. In: *IEEE International Geoscience and Remote Sensing Symposium, 2006*, Denver, USA, pp.168-171.

ACKNOWLEDGEMENTS

The research was funded through a Swedish National Space Board grant. The RADARSAT-2 Fine-beam QuadPol SAR images were provided by the Canadian Space Agency through RADARSAT-2 Data for Research Use Program.

FUSION OF OPTICAL AND RADAR REMOTE SENSING DATA: MUNICH CITY EXAMPLE

G. Palubinskas *, P. Reinartz

German Aerospace Center DLR, 82234 Wessling, Germany - (gintautas.palubinskas, peter.reinartz)@dlr.de

KEY WORDS: Fusion, Imagery, Multisensor, SAR, Optical, Orthoimage, Acquisition, Geometry

ABSTRACT:

Fusion of optical and radar remote sensing data is becoming an actual topic recently in various application areas though the results are not always satisfactory. In this paper we analyze some disturbing aspects of fusing orthoimages from sensors having different acquisition geometries. These aspects are errors in DEM used for image orthorectification and existence of 3D objects in the scene. We analyze how these effects influence a ground displacement in orthoimages produced from optical and radar data. Further, we propose a sensor formation with acquisition geometry parameters which allows to minimize or compensate for ground displacements in different orthoimages due the above mentioned effects and to produce good prerequisites for the following fusion for specific application areas e.g. matching, filling data gaps, classification etc. To demonstrate the potential of the proposed approach two pairs of optical-radar data were acquired over the urban area – Munich city, Germany. The first collection of WorldView-1 and TerraSAR-X data followed the proposed recommendations for acquisition geometry parameters, whereas the second collection of IKONOS and TerraSAR-X data was acquired with accidental parameters. The experiment fully confirmed our ideas. Moreover, it opens new possibilities for optical and radar image fusion.

1. INTRODUCTION

Data fusion is an extremely emerging topic in various application areas during the last decades. Image fusion in remote sensing is one of them. However fusion of different sensor data such as optical and radar imagery is still a challenge. In this paper the term ‘radar’ is equivalent to Synthetic Aperture Radar (SAR). Though the data fusion is well spread over different communities there are quite few attempts of its definition. The first one is the so called JDL information fusion definition (U.S., 1991) popular in military community. This definition is based on the functional model including processing levels and full control on sensors thus making it difficult to transfer to other communities. Another data fusion definition more suitable for a broader community is introduced in (Pohl, 1998) mainly emphasizing (and thus simultaneously limiting to) methods, tools and algorithms used. A more general definition is proposed in (Wald, 1999; Data Fusion Server) as a formal framework in which are expressed the means and tools for the alliance of data originating from different sources. According this definition an alignment of information originating from different sources now becomes a part of the fusion process itself.

There exist numerous remote sensing applications e.g. image matching and co-registration (Suri, 2008), pan sharpening (Klonus, 2008), orthoimage generation, digital elevation model (DEM) generation, filling data gaps, object detection, recognition (Soergel, 2008), reconstruction (Wegner, 2009) and classification (Palubinskas, 2008), change detection, etc which are already profiting or can profit significantly from a data fusion.

For the fusion of data from sensors exhibiting different acquisition geometries such as optical and radar missions it is important to understand their influence on a fusion process and to optimize it if necessary. Of course having not a full control on sensors as in a military community it is not so easy but is still possible to influence some acquisition parameters. In this paper we analyze the effect of ground displacements in orthoimages of optical and radar sensors due to the height error in the DEM used during orthorectification process and 3D objects characteristics (height) for various data acquisition parameters such as sensor look angle (elevation) and look direction, satellite flight direction and sun illumination direction.

The paper is organized in the following way. First, the methodology used for the proposed approach is presented in detail. Then, data used in experiments are described, followed by the presentation of experimental results, conclusion, acknowledgments and finally references.

2. METHOD

In this section we analyze two effects: height error in DEM used during orthorectification process and 3D object height and their influence on ground displacements in orthoimages from optical and radar sensors. The study results in a proposal of several data acquisition parameters: sensor look angle (elevation) and look direction, satellite flight direction and sun illumination direction leading to an optimal sensor formation for the following optical and radar data fusion.

2.1 Effect of DEM height

Ground displacement Δx due the height error Δh in the DEM for an optical and a radar sensor orthoimage is shown in Figure 1.

* Corresponding author.

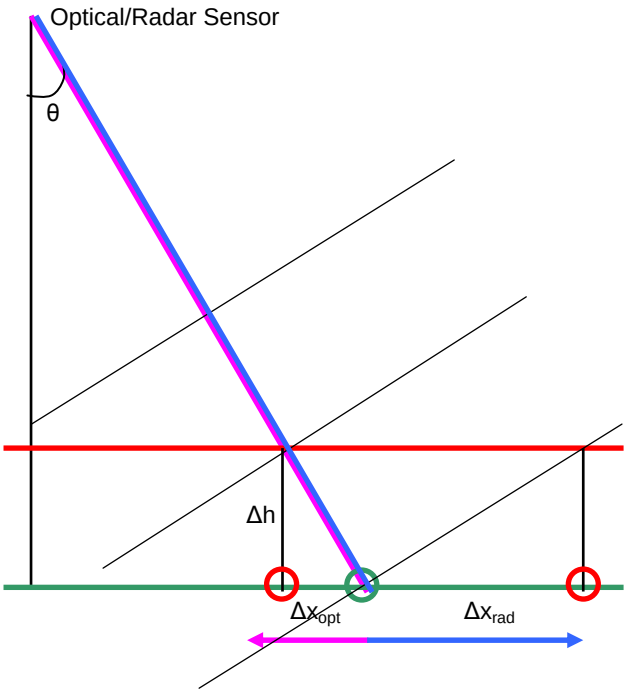


Figure 1. Ground displacement Δx due the height error Δh (positive and negative) in a flat DEM for an optical and radar sensor orthoimage. Look directions: pink line for an optical sensor, blue line – radar sensor. The green horizontal line stands for a true DEM, whereas the red line stands for an error in the DEM (same for both sensors). Similarly, the green circle stands for a true ground position of a 2D point, whereas the red circle – a displaced position. Thin black lines perpendicular to blue line show approximately the radar wave propagation. Flight track is into plane.

Ground displacements are equal to

$$\Delta x_{opt} = \Delta h \cdot \tan\theta_{opt} \quad (1)$$

for optical sensors and

$$\Delta x_{rad} = \frac{\Delta h}{\tan\theta_{rad}} \quad (2)$$

for radar sensors. We have to note, that ground displacements are towards the sensor for the optical case and opposite for the radar case (sign of displacement is ignored in formulae). For details on radar geometry see e.g. (Oliver, 1998).

2.2 Effect of 3D object height

Ground displacement Δx for a 3D object of Δh height for an optical and a radar sensor orthoimage is shown in Figure 2.

Formulae for ground displacements are the same as in the previous sub-section: for optical case equation (1) and radar case - (2). The only difference is a displacement direction: it is away from sensor for the optical case and opposite for the radar case.

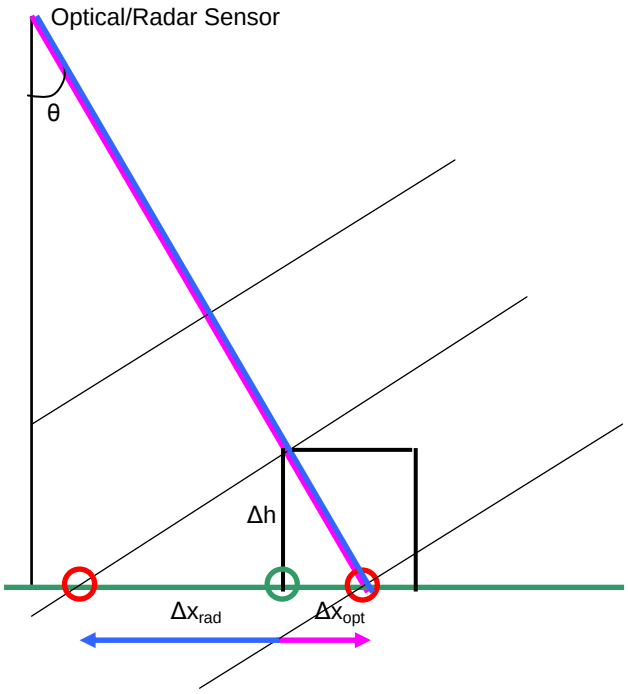


Figure 2. Ground displacement Δx for a 3D object of Δh height for an optical and radar sensor orthoimage. Look directions: pink line for optical sensor, blue line – radar sensor. The green horizontal line stands for a flat DEM, which doesn't include height information of objects. The green circle stands for a true ground position of a 3D point, whereas the red circle – a displaced position. Thin black line perpendicular to blue line shows approximately the radar wave propagation. Flight track is into plane.

2.3 Equality of displacements

We have seen in the previous sub-sections that sizes of ground displacement are different (different formulae) for optical and radar sensors and, moreover, displacement directions are opposite for different sensors. The size equality of ground displacements

$$\Delta x_{opt} = \Delta x_{rad} \quad (3)$$

is fulfilled for the following sensor look (elevation) angles

$$\theta_{opt} + \theta_{rad} = 90° \quad (4)$$

We have to note, that smaller ground displacements are obtained in case of

$$\theta_{opt} < \theta_{rad} \quad (5)$$

In order to compensate opposite displacement directions for different sensors the look directions of different sensors should be opposite. Under the conditions of (4) or (5) structures in optical and radar images appear almost in the same positions thus leading to an easier interpretation and further processing of joint data.

2.4 Optimal sensor constellation

In this sub-section we propose an optimal optical and radar sensor formation for an image acquisition compensating/ minimizing ground displacement effects of different sensors (see Figures 3, 4). A sum of look angles should give approximately 90° (Figure 3).

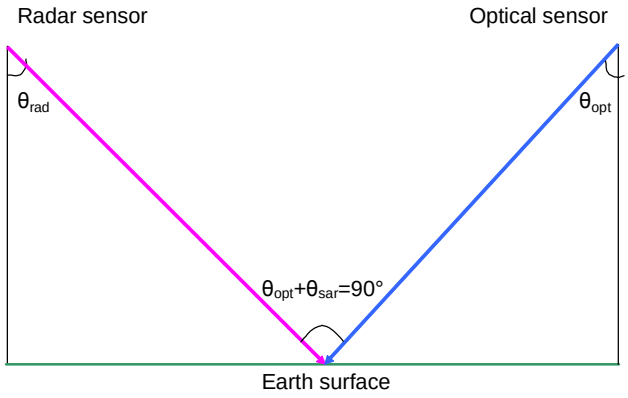


Figure 3. Proposed optical and radar sensor formation is illustrated. A sum of look angles should give 90°.

Flight directions should be as parallel as possible and perpendicular to look directions which are opposite for different sensors (Figure 4). Same flight directions are not required in general e.g. airborne case. A sun illumination direction is from an optical sensor to the target on the Earth in order to see a side of a 3D object which is in shadow in radar image and thus enable full reconstruction of a 3D object. This sensor configuration allows a recovery of 3D object shadows during further data fusion, except a case when the Sun illumination direction is the same as for SAR look direction. Displayed left looking radar and right looking optical sensor formation can be preferable due to the Sun illumination direction which is from an optical sensor to the target on the Earth in order to see that side of a 3D object which is in shadow in the radar image and thus enable full reconstruction of a 3D object. Of course, the second sensor formation with a right looking radar and left looking optical sensor can be useful for data fusion too.

Our approach could be applied in both airborne and space remote sensing. As an example we consider the latter one.

Currently, most space optical remote sensing satellites are acquiring data in descending mode, so a radar satellite should also acquire in a descending orbit. Thus both satellites would fly in the same direction (quasi-parallel orbits). The requirement of opposite look angles and a special sun illumination direction result in a left looking radar sensor and a right looking optical sensor what is achievable with current radar missions though not in a nominal mode (left looking radar). Additionally, larger look angle of SAR sensor than look angle of optical sensor allows minimizing the sizes of ground displacements.

3. DATA

The German Aerospace Center DLR and DigitalGlobe have been engaged in a modest R&D project to investigate complementary uses of Optical and Radar data. Coordinated collections of high resolution TerraSAR-X (TS-X) and WorldView-1 (WV-1) data during July-August 2009 have been acquired. For this experiment one scene of WorldView-1 over Munich city, Germany has been acquired. For more detail on TS-X see (Eineder, 2005). Other scenes of the same urban area of TerraSAR-X and IKONOS have been ordered from existing archives.

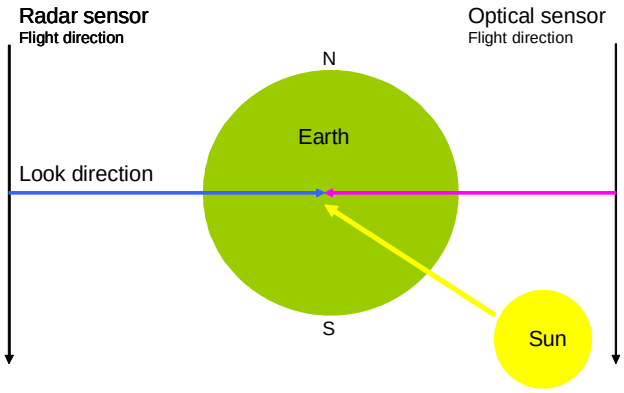


Figure 4. Proposed optical and radar sensor formation is illustrated. Flight directions should be parallel, in same direction and perpendicular to look directions which are opposite for different sensors (right drawing). Sun illumination direction is from an optical sensor to the target on the Earth.

4. EXPERIMENTS

Two experiments, one with a proposed sensor formation and one with an accidental sensor formation were performed to show the potential of our approach. The optical image has been corrected for absolute position by ground control, which yielded a global shift value of approximately 10 m in x-direction for the WV-1 data and 6 m in x-direction and 2 m in y-direction for the IKONOS data in comparison to image rectification without ground control. TS-X data Enhanced Ellipsoid Corrected (EEC) product can be used without ground control, since absolute positioning Root Mean Square Error (RMSE) for the Spotlight mode is in the order of 1 m (Bresnahan, 2009).

4.1 Proposed sensor formation

Scene parameters for the proposed sensor formation experiment are presented in Table 5.

Sensor / Parameter	TS-X	WV-1
Image data	7-Jun-2008	18-Aug-2009
Image time (UTC)	05:17:48	10:50:42
Mode	Spotlight HS	PAN
Look angle	49.45° Right	38.3° Left
Polarization	VV	-
Product	EEC	L2A
Resolution gr x az (m)	1.0 x 1.14	0.89 x 0.65

Table 5. Scene parameters of the first experiment over Munich city

Part of Munich city acquired by WV-1 (upper image) and TS-X (lower image) using the proposed satellite formation is shown in Figure 7. Yellow grid lines are for better orientation between the two images. Ground objects like streets and plazas as well as structures e.g. buildings and trees can be easily detected in both images and are found at the same geometrical position in both images. Only the feet of the buildings, which are differently projected in the radar image due to foreshortening

are found at different positions. The roofs and tree crowns are well in place and can be overlayed correctly for any further processing. Groups consisting of 2, 5 and 6 buildings are highlighted in blue color in both images to show a good correspondence.

4.2 Accidental sensor formation

Scene parameters for the accidental sensor formation experiment are presented in Table 6.

Sensor / Parameter	TS-X	IKONOS
Image data	25-Feb-2008	15-Jul-2005
Image time (UTC)	16:51:15	10:28:06
Mode	Spotlight HS	PAN
Look angle	22.75° Right	5.0° Right
Polarization	VV	-
Product	EEC	Orthoimage
Resolution gr x az (m)	1.6 x 1.3	0.8 x 0.8

Table 6. Scene parameters of the second experiment over the city of Munich

Again, part of Munich city acquired by IKONOS (upper image) and TS-X (lower image) using the accidental satellite formation is shown in Figure 8. Yellow grid lines are for better orientation between two images. For this case it is quite difficult to find corresponding structures in the two images. Only ground objects like streets can be found at similar places but buildings are represented in very different geometry and can be hardly allocated to each other. Also from a radiometric point of view the differences are higher than in Figure 7 probably due to different shadow properties. The same groups consisting of 2, 5 and 6 buildings as in sub-section 4.1 are highlighted in blue color in both images again. In this case it is quite difficult to identify the same number of buildings in both images.

5. CONCLUSIONS

In this paper we address a problem of fusion of optical and radar remote sensing imagery. Alignment of information coming from different sources is an important prerequisite for the following fusion in various applications. Especially for a rapid fusion of optical and radar data a specific imaging is of advantage. We propose an optical and radar sensor formation which accounts for different acquisition geometries and minimizes displacements for ground and 3D-objects in orthoimages of optical and radar sensors. The preferred sensor formation is a perpendicular viewing from the two sensor systems due to the complimentary nature of their viewing geometries. For this case the image geometries are nearly independent to errors in the underlying DEM and especially to buildings or other 3D objects, not represented in the DEM. A fast and consistent overlay of the two data sets for on ground and other surfaces is reached. As an example two pairs of high resolution optical (WorldView-1 and IKONOS) and radar (TerraSAR-X) images have been acquired over the urban area - Munich city in Germany – for different sensor formations. Results show a great potential of the proposed approach for further applications of data fusion with optical and radar instrumentation since the geometric positions of the objects can be observed at the same absolute position.

REFERENCES

Bresnahan, P., 2009. Absolute Geolocation Accuracy Evaluation of TerraSAR-X Spotlight and Stripmap Imagery – Study Results. In: *Proceedings of Civil Commercial Imagery Evaluation Workshop, 31 March – 2 April 2009, USGS, Fairfax Virginia, USA.*

Eineder, M., Schättler, B., Breit, H., Fritz, T. and Roth, A., 2005. TerraSAR-X SAR products and processing algorithms. In: *Proc. of IEEE International Geoscience and Remote Sensing Symposium (IGARSS'05), 25-29 July, 2005, Seoul, Korea*, IEEE, vol. VII, pp. 4870-4873.

Klonus, S., 2008. Comparison of Pansharpening Algorithms for Combining RADAR and Multispectral Data. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.* Volume XXXVII, Part B6b, Beijing, China, pp. 189-194.

Palubinskas, G. and Datcu, M., 2008. Information fusion approach for the data classification: an example for ERS-1/2 InSAR data. *International Journal of Remote Sensing*, vol. 29(16), pp. 4689-4703.

Pohl, C. and van Genderen, J. L., 1998. Multisensor image fusion in remote sensing: concepts, methods and applications. *International Journal of Remote Sensing*, 19(5), pp. 823-854.

Oliver, C. and Quegan, S., 1998. *Understanding Synthetic Aperture Radar Images.* Artech House, Boston.

Suri, S. and Reinartz, P., 2008. Application of Generalized Partial Volume Estimation for Mutual Information based Registration of High Resolution SAR and Optical Imagery. In: *Proc. of 11th International Conference on Information Fusion (FUSION'2008), June 30 – July 3, 2008, Cologne, Germany*, pp. 1257-1264.

The Data Fusion Server. Available from: http://www.data-fusion.org/ (Accessed 24 February 2010)

U.S. Department of Defense, Data Fusion Subpanel of the Joint Directors of Laboratories (JDL), Technical Panel for C3, "Data fusion lexicon," 1991.

Wald, L., 1999. Some terms of reference in data fusion. *IEEE Transactions on Geosciences and Remote Sensing*, 37, 3, pp. 1190-1193.

Wegner, J.D., Auer, S., and Soergel, U., 2009. Accuracy Assessment of Building Height Estimation from a High Resolution Optical Image Combined with a Simulated SAR Image. In: *Proc. of ISPRS Hannover Workshop 2009 - High Resolution Earth Imaging for Geospatial Information, 2-5 June, 2009, Hannover, Germany*, ISPRS, vol. XXXVIII-1-4-7, part W5.

ACKNOWLEDGEMENTS

We would like to thank DigitalGlobe for the collection and provision of WorldView-1 scene over Munich city,and European Space Imaging for providing the IKONOS-2 scene.

Figure 7. Part of Munich city acquired by VW-1 (upper image) and TS-X (lower image) using the proposed satellite formation. Yellow grid lines are for better orientation between two images. Red arrows show flight (az) and look (rg) directions.

Figure 8. Part of Munich city acquired by IKONOS (upper image) and TS-X (lower image) using the accidental satellite formation. Yellow grid lines are for better orientation between two images. Red arrows show flight (az) and look (rg) directions.

TIDAL WETLAND MONITORING USING POLARIMETRIC SYNTHETIC APERTURE RADAR

S. -E. Park [a, *], D. Kim [b], H.-S. Lee [c], W. M. Moon [d], W. Wagner [e]

[a] Graduate School of Science and Technology, Niigata University, Niigata 950-2181, Japan - s.park@wave.ie.niigata-u.ac.jp
[b] School of Earth and Environmental Sciences, Seoul National University, Seoul, 151-747, Korea - djkim@snu.ac.kr
[c] Department of Civil and Environmental Engineering, Sunchon National University, Chonnam 540-742, Korea - hslee@sunchon.ac.kr
[d] Geophysics, University of Manitoba, Winnipeg, MB R3T 2N2 Canada - wmoon@cc.umanitoba.ca
[e] IPF, Vienna University of Technology, Gusshausstr. 27–29, 1040 Vienna, Austria - ww@ipf.tuwien.ac.at

KEY WORDS: Synthetic Aperture Radar, Tidal Wetlands, Radar Polarimetry, Microwave Scattering from Rough Surface

ABSTRACT:

Tidal wetlands including intertidal flats are highly productive and have dynamic and diverse ecosystems. Despite the importance of the tidal flats and associated coastal habitats, theses areas are at risk due to high development pressure, such as reclamation and marine pollution. Because of their poor accessibility, remote sensing techniques are the most effective tool for tidal flat observation. Particularly, microwave remote sensing using synthetic aperture radar (SAR) system has great potential for quantitative monitoring and mapping of coastal wetlands. This study aims to review and develop effective methods of extracting geophysical information of tidal wetlands. Fully polarimetric forward/inverse scattering models have been developed for quantitative estimation of geophysical parameters. This study aims to review and develop effective methods of extracting geophysical information of intertidal mudflats including surface geometric characteristics, such as the roughness of the scattering surface, from polarimetric SAR data. In addition, an extension of previous study to fully polarimetric space-borne SAR data sets is presented in this paper.

1. INTRODUCTION

Coastal wetlands including tidal flats are the zone of interaction between marine and terrestrial environments. They have dynamic and diverse ecosystems and provide highly productive fishery areas. Despite the importance of the tidal flats and associated coastal habitats, these areas are at risk due to high development pressure, such as reclamation and marine pollution. In addition, coastal wetlands are highly vulnerable to climate changes. Because of their poor accessibility both from sea and land, monitoring and mapping of tidal flat environments from in situ measurements in field are very difficult.

Remote sensing can provide large spatial coverage and non-intrusive measurement over the Earth's surface. Previous studies have focused on the use of optical sensors for remote sensing of tidal flats. Because of the repetitive tidal event and dynamic sedimentary process, however, integrated observation of tidal flats from multi-sourced data sets is essential for mapping and monitoring tidal flats.

Microwave remote sensing using synthetic aperture radar (SAR) system can be a complementary tool for tidal flat observation especially due to their high spatial resolution and all-weather imaging capability. Recently several studies have reported on investigation of tidal flats using single polarization SAR data [Van der Wal et al., 2005], multi-frequency approach [Gade et al., 2008], and dual or full-polarimetric approach [Park et al., 2009].

This study aims to review and develop effective methods of extracting geophysical information of intertidal mudflats including surface geometric characteristics, such as the roughness of the scattering surface, from polarimetric SAR data. Roughness of the surface sediments in intertidal flats represents both biogenic and physical depositional characteristics of sediments [Reineck and Singh, 1980]. In addition, it can also be used for describing the land-use characteristics in intertidal flats such as fishery activities. In Section II, previous researches on remote sensing of tidal wetlands are reviewed, and the inversion algorithms of surface roughness parameters are presented in Section III. Experimental results on the roughness estimation of intertidal flats are discussed in Section IV. Finally, summary and concluding remarks are presented in Section V.

2. REVIEW OF REMOTE SENSING OF TIDAL FLAT

2.1 Tidal flat monitoring from optical sensors

Remote sensing using optical sensors (e.g., Landsat TM) have been applied for mapping of surface sediment distributions of tidal flats [Yates et al., 1993; Rainey et al., 2000, Ryu et al., 2004]. The distribution of sediments of a given particle size is of great interest in the field of morpho-dynamics and sedimentary process of coastal environments.

Yates et al. (1993) tested several classification algorithms to distinguish between muddy and sandy flats using Landsat TM data. In their study, muddy and sandy flats are determined based on the sediment 'critical' grain size of 0.063 mm. Results indicated that all the classification methods showed a better performance for muddy flats than for sandy flats.

Ryu et al. (2004) claimed to use critical grain size of 0.25 mm for practical classification of fine and coarse sediments. In addition, they investigated the effect of water contents and topography which affect optical reflectance values on the classification of surface sediment distribution. Particularly,

* Corresponding author.

water content which can be remained on the intertidal flat surface for considerable times after exposure, significantly affect the spectral response of the sediment.
The classification performance can be improved by applying coarse critical grain size. Nonetheless classification of the intertidal surface sediments can still be considered to be in an experimental stage because of poor knowledge on coupled effect of the topography and the water content as well as the grain size of the sediments on the optical reflectance of intertidal sediments.

2.2 Tidal flat monitoring from microwave sensors

Improving temporal resolution is an essential issue of remote sensing of coastal environments due to the dynamic changes in morphological and sedimentary processes of tidal wetlands. Unlike optical and infrared imaging sensors SAR is an active microwave imaging system for studying Earth's environment with a high spatial resolution. Unlike optical and infrared imaging sensors which rely on reflected or radiated solar energy, imaging radars are more flexible and robust in their tidal flat observation capabilities because they are independent of sunlight and weather conditions.
Theoretical relationships between backscatter coefficients and surface geophysical parameters have been well developed over several decades. In general, radar signals backscattered from bare soil surface can be expressed by the dielectric properties of material and statistical roughness characteristics of scattering surfaces. In case of the tidal flat application, however, one can assume that surface sediments are fully saturated by sea water particularly in mud flats which dominates the tidal flats around Korean peninsular. Consequently, the effect of dielectric constants on backscattered signals can be neglected in the specific case of the intertidal mudflats [Van der Wal et al., 2005; Gade et al., 2008; Park et al., 2009]. Several studies have been proposed the inversion algorithm of remained unknown geophysical parameters, such as the rms height, *s*, and correlation length, *l*, of surface sediments, from radar measurements.
In Van der Wal et al (2005), the rms height of surface sediment was obtained by an empirical regression approach based on the C-band, VV-polarized backscattering coefficient of ERS SAR data. Due to its simplicity, it can be applied to the common SAR systems which operate in single-frequency and single-polarization. However, practical use of this approach for parameter retrieval needs time-consuming calibration work on empirical relationships for various radar configurations and surface conditions.
On the other hand, Gade et al. (2008) proposed another roughness retrieval algorithm based on multi-frequency SAR data. The rms height and the correlation length were obtained from VV-polarized backscattering coefficients of X-, C-, and L-band SAR data. In this case, the retrieval algorithm is based on the theoretical scattering model, such as Integral Equation Method (IEM) [Fung, 1992]. Since the IEM model has a broader range of validity than classical scattering models, e.g., Kirchhoff approach and the small perturbation method, it can be used for roughness retrieval of diverse surface sediments. However, it has an inherent limitation in operational use of space-borne SAR data to monitor tidal flats, because of a high temporal variety of tidal flat environment and a rare availability of multi-frequency SAR data acquired simultaneously.
Recently, Park et al. (2009) proposed an alternative roughness retrieval algorithm based on fully polarimetric SAR data. This roughness inversion technique has been validated using L-band NASA/JPL AIRSAR data sets. Roughness parameters estimated from SAR data are in reasonably good agreements with those from in-situ measurements. This study aims to present an extension of previous study to polarimetric space-borne SAR data.

3. SAR POLARIMETRY OF TIDAL FLAT

3.1 Polarimetric SAR remote sensing

The microwave transmitted by a radar system is characterized by its frequency and its polarization state. Conventional SAR systems operate with a single fixed polarization antenna for both transmission and reception of microwave frequency signals as shown in Table 1. Today, there is a rapidly increasing interest in the application of radar polarimetry for Earth observation due to increasing availability of polarimetric space-borne radar sensors, such as ALOS-PALSAR, RADARSAT-2, and Terra-SAR-X and follow-up satellite sensor systems.
The backscattered wave by the target is described in either the incident components $\vec{E}^i$ or by the scattered components $\vec{E}^s$, which are related by the complex 2×2 scattering matrix [*S*] of the target, defined according to

$$\begin{bmatrix} E_H^s \\ E_V^s \end{bmatrix} = \frac{e^{-jkr}}{r} \begin{bmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{bmatrix} \begin{bmatrix} E_H^i \\ E_V^i \end{bmatrix}, \text{ or } \vec{E}^s = \frac{e^{-jkr}}{r}[S]\vec{E}^i \quad (1)$$

where k is wave number, r is the distance between the target and the receiving antenna. The element $S_{pq} = |S_{pq}| \exp(j\varphi_{pq})$ is dependent on the characteristic of the target and on the direction of incident and scattered fields.
There is an advantage in many applications to expressing target scattering properties in terms of a complex feature vectors, comprising three elements in the monostatic backscatter case. The lexicographic feature vectors in linear and circular basis $\vec{k}_L$ and $\vec{k}_{RL}$, and Pauli feature vector $\vec{k}_P$ are defined as:

$$\vec{k}_L = [S_{HH} \quad \sqrt{2}S_{HV} \quad S_{VV}]^T ; \; \vec{k}_{RL} = [S_{RR} \quad \sqrt{2}S_{RL} \quad S_{LL}]^T ;$$
$$\vec{k}_P = \tfrac{1}{\sqrt{2}}[S_{HH} + S_{VV} \quad S_{HH} - S_{VV} \quad 2S_{HV}]^T . \quad (2)$$

One of the most important properties of radar polarimetry is the fact that once a target response is acquired in a polarization basis, the response in any basis can be obtained from a simple transformation without any additional measurements. All of theses target feature vectors are related with each others by

Satellite, Instrument	Frequency	Polarization
ERS-1, SAR (1991-2000)	C-band	VV
JERS-1, SAR (1992-1998)	L-band	HH
ERS-2, SAR (1995~)	C-band	VV
RADARSAT-1 (1995~)	C-band	HH
ENVISAT, ASAR (2002~)	C-band	Twin
ALOS, PALSAR (2006~)	L-band	Quad
RADARSAT-2 (2007~)	C-band	Quad
TerraSAR-X (2007~)	X-band	Dual
COSMO-SkyMed (2007~)	X-band	Dual
RISAT (2009)	C-band	Quad
HJ-1-C (2010)	S-band	Quad
SAOCOM (2010)	L-band	Quad
Sentinel-1 (2011)	C-band	Dual

Table 1. Current and future space-borne SAR systems

unitary transformation matrices [Boerner et al., 1998]. Consequently, the amount of information about a given scatterer can be increased, allowing a better characterization of scattering properties. An advantage of the Pauli basis equation (2) is that [*S*] is projected on to orthogonal basis matrices that represent simple scattering mechanisms. The first component $S_{HH}+S_{VV}$ dominates in single-bounce surface scatter, while the second component $S_{HH}-S_{VV}$ dominates in double-bounce scatter. The cross-pol element S_{HV} will be strong for backscatter from depolarizing media.

3.2 Roughness retrieval from polarimetric SAR

Since effect of dielectric constant on radar measurements can be neglected in tidal flats, the set of roughness parameters {*s*, *l*} can be can be obtained simultaneously from two independent polarization measurements. In particular, increased measurement sensitivity to surface roughness has been reported by using the circular-polarization coherence $|\rho_{RRLL}|$ defined as:

$$|\rho_{RRLL}| = |<S_{RR}S_{LL}^{*}>| \Big/ \sqrt{<S_{RR}S_{RR}^{*}><S_{LL}S_{LL}^{*}>} \qquad (3)$$

Consequently, the pair of $|\rho_{RRLL}|$ and the co-pol response of the coherency matrix $|S_{HH}+S_{VV}|^2$ showed better performance in the roughness parameter retrieval than conventional co-and cross-polarized backscattered coefficients. Therefore, roughness parameters of the target surface can be estimated by the minimization procedure:

$$\min\left\{\text{norm}\left\{\begin{bmatrix}|\rho_{RRLL}|\\ |S_{HH}+S_{VV}|^2\end{bmatrix}-F\left(\begin{bmatrix}s\\ l\end{bmatrix}\right)\right\}\right\}. \qquad (4)$$

The extended-Bragg model [Schuler et al., 2002; Hajnsek et al., 2003] has been used for the scattering model *F* in (4). This roughness inversion technique has been validated using L-band NASA/JPL AIRSAR data sets as shown in Figure 2. Roughness parameters relative to wavelength, *ks* and *kl*, derived from SAR data are in reasonably good agreements with those from in-situ

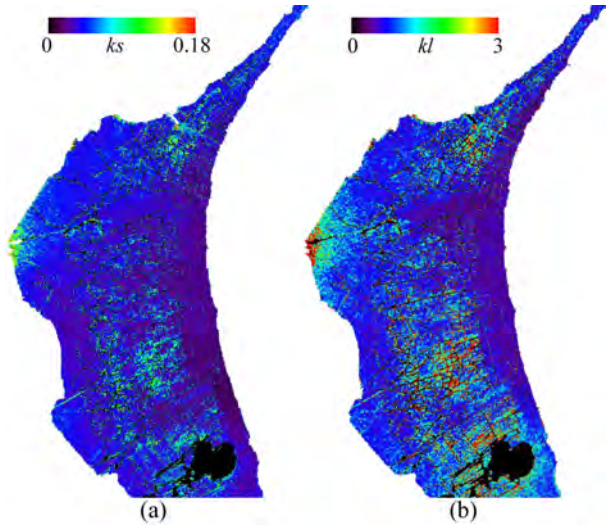


Figure 2. Radar derived (a) rms height and (b) correlation length [Park et al., 2009].

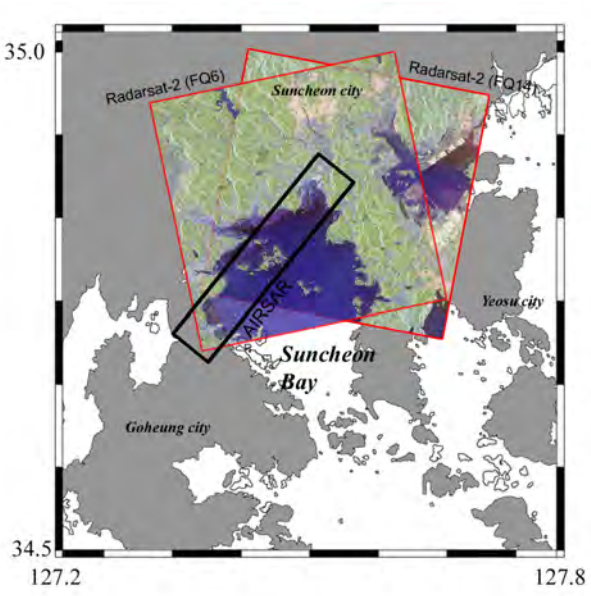


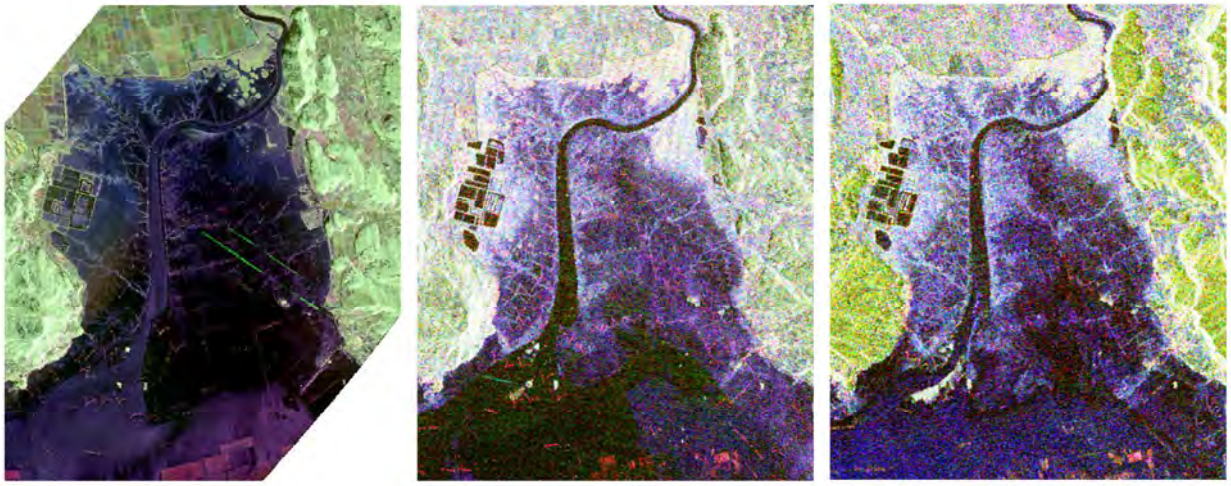
Figure 3. Suncheon Bay study area.

measurements. More details on the performance analysis of roughness retrieval algorithm can be found in [Park et al., 2009]. In case of space-borne SAR remote sensing, however, the surface of tidal flat sediment is generally very smooth in L-band frequency resulting in low backscattering signal relatively close to the radar noise floor. Therefore, higher frequencies, such as C- and X-band SAR data could be more appropriate to recognize fine details than L-band.

4. ESTIMATION OF SURFACE ROUGHNESS PARAMETERS

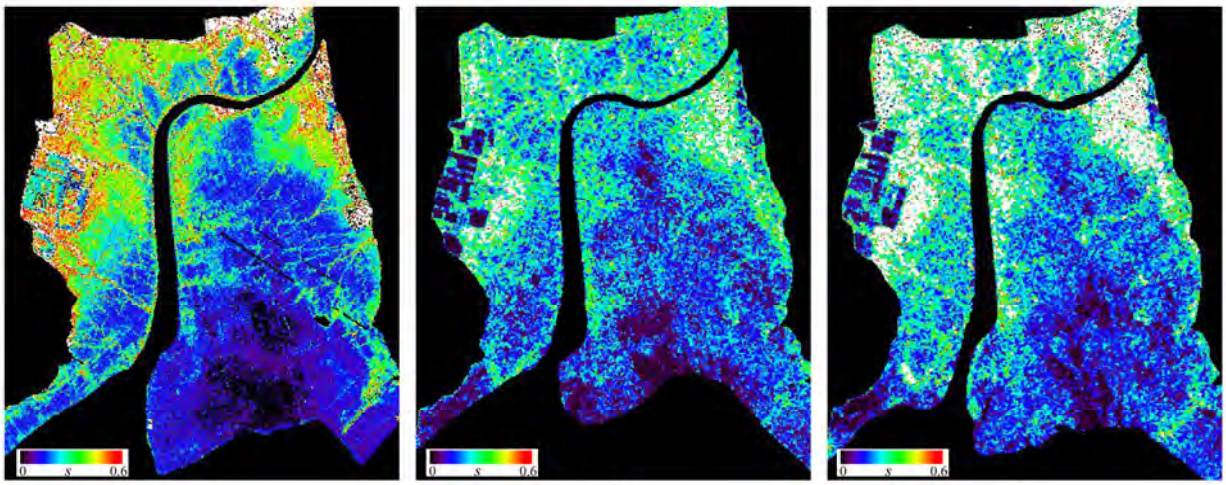
The south and west coastal zones of Korean peninsula are well known for their large tidal ranges and vast expanses of intertidal flats. Suncheon Bay study area in the southern coast of the Korean peninsular is one of the highly productive fishery region and provides habitat of various fish and shellfishes as well as migration birds. Recently, Suncheon Bay tidal wetland was registered as a RAMSAR Site for international convention of wetlands conservation.

Figure 3 shows available polarimetric SAR data sets over Suncheon Bay study site. The fully polarimetric NASA/JPL AIRSAR data were acquired at L-band during PACRIM-II Korea campaign on September 30th, 2000. In addition, two sets of RADARSAT-2 fully polarimetric data were obtained over the Suncheon Bay study area. The first data set was acquired on November 4, 2008 at fine beam mode (FQ14) of descending orbit. The radar incidence angles vary between 33.5° and 35.1°. The second one was acquired on December 14, 2008. In this case, the image was obtained in ascending orbit at FQ6 beam mode in which the radar incidence angles span 24.5° to 26.4°. The first data (FQ14) was acquired at an ebbing tide with the tidal elevation of 105 m, while the second one was acquired on the beginning of the flowing tide with the tidal elevation of 110 m. Figure 4 shows AIRSAR and RADARSAT-2 images for the specific test site near the mouth of the river flows into the Bay. New artificial structures can be identified in the western part of the tidal flats. Compared with AIRSAR, C-band RADARSAT-2 backscattering coefficients show higher sensitivity to vegetation and sediment structures in tidal flats.



(a) (b) (c)

Figure 4. SAR images of Suncheon Bay study area: (a) L-band AIRSAR (Sep. 2000), (b) RADARSAT FQ14 (Nov. 2008), and (c) RADARSAT FQ6 (Dec. 2008)



(a) (b) (c)

Figure 5. Radar derived rms height maps (in cm) of the tidal flats as a result of the inversion of (a) L-band AIRSAR, (b) RADARSAT-2 FQ14 beam mode, and (c) RADARSAT-2 FQ6 beam mode data sets.

Figure 5 shows the results of roughness parameter retrievals from polarimetric descriptors $|\rho_{RRLL}|$ and $|S_{HH}+S_{VV}|^2$. Pixels correspond to the land and the ocean were excluded from the inversion process. The rms height derived from NASA/JPL AIR SAR data acquired on September 30, 2000 is shown in Figure 5(a). It can infer the performance of the roughness retrieval from RADARSAT-2 data despite of eight years difference between AIRSAR and RADARSAT-2. The rms heights derived from C-band RADARSAT-2 data show good agreements with those from AIRSAR data. Consequently, the fully polarimetric approaches which have been validated originally on the basis of L-band air-borne SAR data are also applicable to the C-band space-borne RADARSAT-2 data for roughness parameter retrievals of surface sediments in tidal flats.

5. CONCLUSION

Despite of different incidence angles and looking directions of FQ14 beam mode in ascending orbit and FQ6 mode in descending orbit, roughness parameters derived from two RADARSAT-2 data are very similar to each other. Therefore, one can reduce the time gap between each observation through a combined use of different beam modes and orbits of RADARSAT-2. However, there are some areas show changes in rms heights during two acquisitions particularly in mudflat near waterline and tidal channel. They reflect the difference of tidal heights, biological activities, and texture, salinity, and moisture contents of surface sediments. Relating radar derived roughness parameters to fundamental environmental processes in tidal flats will be further investigated through time series of polarimetric SAR data sets and detailed in-site measurements.

REFERENCES

Boerner, W.-M., H. Mott, E. Luneburg, C. Livingstone, B. Brisco, R. J. Brown, and J. S. Paterson, 1998, *Polarimetry in Radar Remote Sensing: Basic and Applied Concepts*, Chapter 5 in F.M. Henderson, and A.J. Lewis, (eds.), Principles and Applications of Imaging Radar, vol. 2 of Manual of Remote Sensing, (ed. R.A. Reyerson), Third Edition, John Willey &

Sons, New York.

Fung, A. K., Z. Li, and K. S. Chen, 1992, Backscattering from a randomly rough surface, *IEEE Trans. Geosci. Remote Sensing*, vol. 30, no. 2, pp. 356-369.

Gade, M., W. Alper, C. Melsheimer, and G. Tanck, 2008, Classification of sediments on exposed tidal flats in the German Bight using multi-frequency radar data, *Remote Sens. Environ.*, vol. 112, pp. 1603–1613.

Hajnsek, I., E. Pottier, and S. R. Cloude, 2003, Inversion of Surface Parameters from Polarimetric SAR, *IEEE Trans. Geosci. Remote Sensing*, vol. 41, no. 4, pp. 727-744.

Park, S.-E., W. M. Moon, D. Kim, and J.-E. Kim, 2009, Estimation of surface roughness parameter in intertidal mudflat using airborne polarimetric SAR data, *IEEE Trans. Geosci. Remote Sensing*, vol. 47, no. 4, pp. 1022-1031.

Rainey, M. P. Rainey, A. N. Tyler, R. G. Bryant, and D. J. Gilvear, 2000, The influence of surface and interstitial moisture on the spectral characteristics of intertidal sediments: implications for airborne image acquisition and processing, *Int. J. Remote Sens.*, vol. 21, no. 16, pp. 3025–3038.

Reineck, H. E., and I. B. Singh, 1980, *Depositional Sedimentary Environments*. Springer-Velag, New York.

Ryu, J. H., Y. H. Na, J. S. Won, and R. Doerffer, 2004, A critical grain size for Landsat ETM+ investigations into intertidal sediments: a case study of the Gomso tidal flats, Korea, *Estuar. Coast. Shelf S.*, vol. 60, no. 3, pp. 491-502.

Schuler, D. L., J. S. Lee, D. Kasilingam, and G. Nesti, 2002, Surface Roughness and Slope Measurements using Polarimetric SAR Data, *IEEE Trans. Geosci. Remote Sensing*, vol. 40, no. 3, pp. 687-698.

Van der Wal, D., P. M. J. Herman, A. W. Van den Dool, 2005, Characterisation of surface roughness and sediment texture of intertidal flats using ERS SAR imagery, *Remote Sens. Environ.*, vol. 98, pp. 96-109.

Yates, M. G., A. R. Jones, S. McGrorty, and J. D. Goss-Custard, 1993, The use of satellite imagery to determine the distribution of intertidal surface sediments of the Wash, England, *Estuar. Coast. Shelf S.*, vol. 36, no. 4, pp. 333–344.

AUTOMATIC DETECTION OF BURIED CHANNEL DEPOSITS FROM DENSE LASER ALTIMETRY DATA

B.M.J. Possel[a], R.C. Lindenbergh[a,*], J.E.A. Storms[b]

[a]Dept. of Remote Sensing, Chair of Optical and Laser Remote Sensing, Delft University of Technology – Boudewijn@Possel.nl, R.C.lindenbergh@tudelft.nl
[b]Dept. of Civil Engineering and Geosciences, Section of Applied Geology, Delft University of Technology – J.E.A.Storms@tudelft.nl
http://www.possel.nl

KEY WORDS: LIDAR, structural classification, shallow subsurface, channel deposits

ABSTRACT:

The formation of the current Rhine-Meuse delta mainly took place during the last 12 000 years. Consecutive avulsions, i.e. sudden changes in the course of river channels, resulted in a complicated pattern of sandy channel deposits, surrounded by peat and clay. Knowledge of this pattern is not only interesting from a geohistorical viewpoint, but is also essential when planning and maintaining constructions like roads and dikes. Traditionally, channel deposits are traced using labor intensive soil drilling. Channel deposits are however also recognizable in the polder landscape by small local elevation changes due to differential compaction. The purpose of this research is to automatically map channel deposits based on a structural analysis of high resolution laser altimetry data. After removing infrastructural elements from the laser data, local feature vectors are built, consisting of the attributes slope, curvature and relative elevation. Using a maximum likelihood classifier, 75 million gridded laser points are divided into two classes: buried channel deposits and other. Results are validated against two data sets, an existing paleographic map and a set of shallow drilling measurements. Validation shows that our method of channel deposit detection is hampered by signal distortion due to human intervention in the traditional polder landscape. Still it is shown that relative young deposits (4 620 to 1 700 years Before Present) can be extracted from the laser altimetry data.

1. INTRODUCTION

During the Holocene (approximately 12 000 years - present), much of the western and central part of the Netherlands was aggrading, as active river systems (Rhine and Meuse) transported sediments from the hinterland to the coastline. In combination with sea level rise this resulted in a Holocene sediment sequence of up to 20 meter thickness. As river channels consist predominantly of sand while the adjacent floodplains were dominated by clay deposition and peat formation, a strong grain-size partitioning occurred. Furthermore, frequent shifts in channel location due to avulsions, resulted in a complex subsurface of clay/peat dominated floodplain deposits laterally and vertically alternating with sand-rich channel areas, (Allen, 1965).

Currently, buried channel deposits are recognizable in the landscape, basically due to a process called topographic inversion. This occurs when floodplain deposits on the sides of the buried channels compact at a higher rate than the channel sand itself. At the surface this results in an area with a higher elevation at the locations of buried channel deposits. Note that the sand remaining from an abandoned channel may not start directly at the surface: channels abandoned relatively long ago may meanwhile have been deeply covered by floodplain deposits. The maximum height differences between the buried channels and the surroundings are in the order of a meter for relative large and young channel deposits. To some extent it holds that the thinner and older, that is, deeper the channel deposits, the smaller also the height difference.

Knowledge of the location of these channel deposits is essential when planning and maintaining large construction works as motorways and dikes, (Munstermann et al., 2008). Abrupt and unidentified changes in the subsurface may lead to unexpected differences in compaction, which may lead again to damaged or uneven road surfaces or even to failing dikes.

Traditionally, mapping of the shallow subsurface of the Rhine-Meuse delta is based on soil drillings. A large effort has been made by Dutch Utrecht University: Based on more than 25 years of field research using over 200 000 manual boreholes a paleogeographic map is composed, (Berendsen and Stouthamer, 2001), see also Fig. 1, right. As the drillings require a large amount of manual labor in the field, not the whole Rhine-Meuse delta has been covered in the same amount of detail. Also necessarily some interpretation and interpolation steps were involved in composing the map, which may have introduced local anomalies.

LIDAR data is being used more and more to reveal and highlight morphological and archaeological structures that are hardly visible. In archaeology, LIDAR data has revealed burial mounts, (Kakiuchi and Chikatsu, 2008), Celtic field systems, (Kooistra and Maas, 2008, Humme et al., 2006), and other earthwork features, (M. Doneus and Jammer, 2008). Spatial scales characteristic for a certain landscape type, like small scale roughness, could be identified by a spectral analysis of airborne laser scanning data, (Perron et al., 2008). Previous studies, (Berendsen and Volleberg, 2007, Munstermann et al., 2008), showed that also buried channel deposits can be visualized using airborne laser data from the AHN (Actueel Hoogtebestand Nederland) archive.

In this research it is considered if it is possible to systematically map channel deposits from second generation, high resolution AHN-2 data. In 2012 for every 50 cm grid point in The Netherlands a height value will be available with a precision of about 5 cm, (AHN, 2000). As a test area the so-called Alblasserwaard is used, a polder of 350 km^2, directly east of Rotterdam. The location of this polder is indicated in the inset in Fig. 5. For this polder, a test data set has been kindly made available by provider

* Corresponding author.

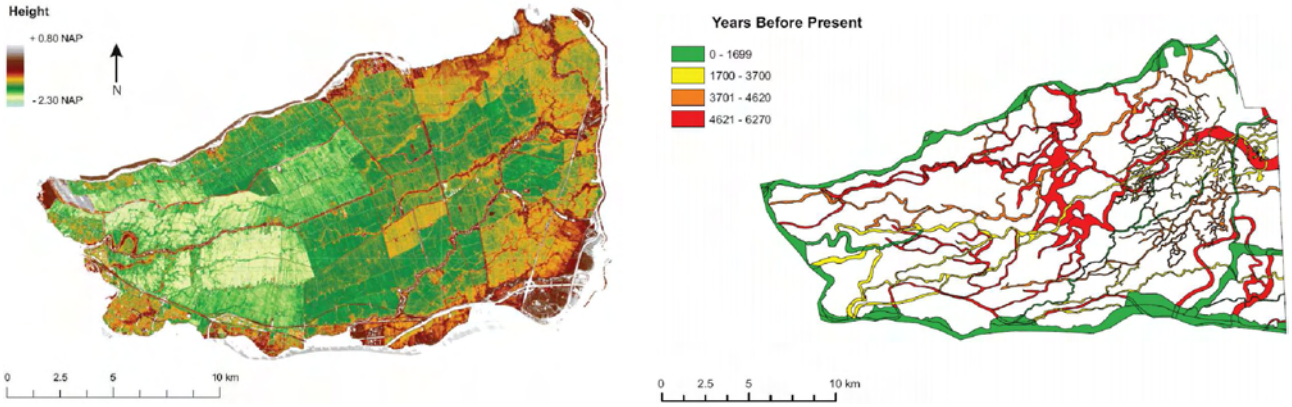


Figure 1: Alblasserwaard. **Left:** Laser altimetry data. **Right:** Paleogeographic map

Fugro Aerial Mapping B.V. and owner Waterboard Rivierenland, consisting of about 1.2 billion gridded points, see Fig. 1, left.

From this data set points representing hard infrastructure are removed in a filter procedure incorporating the Dutch topographic base map GBKN. Remaining points are classified according to four structural attributes into two classes, channel deposit and non channel deposit. In Section 2. this data filtering and classification procedure is discussed. Results are validated in Section 3. against the digital paleogeographic map and against an interpretation of drillings from the Dutch geological database DINOLoket.

2. LIDAR DATA FILTERING AND CLASSIFICATION

In this section methodology is described aiming at the classification of airborne laser altimetry points into two classes, buried channel and non-buried channel deposits. A main challenge in this research is the huge amount of input points. As the input data strongly influences the methodology, these are described first. Then it is described how laser points representing hard infrastructure are removed before describing the actual classification method.

2.1 Data description

For this research FLI-MAP400 VS laser altimetry data is used, measured by Fugro Aerial Mapping BV for the Waterboard Rivierenland. An overview of the entire data set is shown in Fig. 1, left. The data was acquired during three days in August 2007, with a minimum point density of 8 points per m^2. The absolute accuracy of a single point is reported to be 3 cm. From this raw data, Fugro derived a Digital Surface Model (DSM) by removing non-terrain points. The DSM points were consecutively resampled to a 0.5m grid using inverse squared distance weighting and organized in tiles of 1.25×1 km. In total the Alblasserwaard data set was divided into 273 of such tiles To decrease computational efforts, the .5m grid was further downsampled to a 2 m grid. As a result, the input data set for this research consists of roughly 75 million points.

2.2 Removing non-field objects

In the gridded FLI-MAP data still objects like roads, trenches, buildings and water surfaces are present. If unaddressed these objects complicate the detection of buried channel deposits. The laser data is filtered in two steps with the purpose of only keeping data representing fields. In the first step, non-field objects are removed using a mask constructed from GBKN data, in the second step remaining unwanted objects are removed, based on a local variability analysis.

GBKN mask. The 'Grootschalige Basiskaart van Nederland' (GBKN) is the Large Scale Standard Map of The Netherlands and is the most detailed and accurate digital topographical database available in the Netherlands, (GBKN, 2009). It is scale-free, but is comparable to paper maps with a scale between 1:100 and 1:5,000. The precision of a point in comparison to another point in the surrounding is better than 28 cm in suburban areas and better than 56 cm in rural areas. The GBKN has a spaghetti-structure: it only contains classified nodes and edges, for instance road sides, water edges and building contours. Therefore the GBKN map of the Alblasserwaard has to be converted to an area map, consisting of classified segments, see Fig. 2, left and middle. This area map will then be applied as a mask to remove those laser points that are in a polygonal segment from an unwanted class, like 'road'.

To create segments, the GBKN lines have to be automatically connected and converted into classified segments. However, there are errors in the database: lines sometimes do not connect exactly or lines intersect without a node. Such situations have to be identified and adapted. Lines in the GBKN that do not exactly connect are attached to the nearest line or node within a certain distance threshold in a snapping procedure. Here a threshold of 10 cm is used. Self intersections without nodes are removed by adding nodes to the intersection points. Around the resulting area mask, an additional buffer of 3m is added to further limit the influence of unwanted objects: for example, ground close to a road is often disturbed, and cannot be considered as representative for the situation in a field.

Despite this filtering method, unwanted features still remain in the LIDAR data, see Fig. 2, right, like small trenches and other objects not (yet) registered in the GBKN database. To further decrease the influence of unwanted features, isolated points and points with a high local variance were additionally removed.

2.3 Channel classification

The points remaining after the removal of non-field objects are classified into two classes by means of structural classification. For this purpose first structural attributes are determined at each grid point. As a result at each grid point a multi variate feature vector is created that can be used as input for a standard remote sensing classification method.

Slope and curvature attributes At each remaining LIDAR point, the four following attributes are determined: slope, curvature, TPI and smoothed TPI. Slope is chosen as an attribute because at both sides of a buried channel, the elevation is increasing with

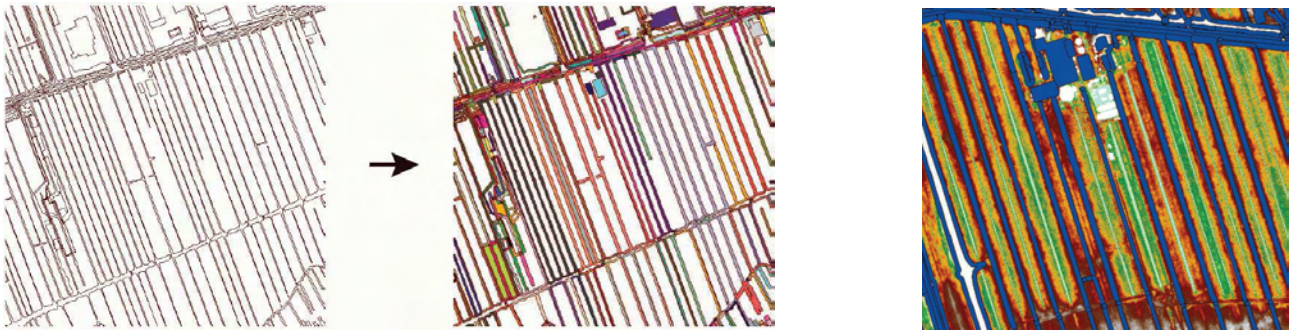


Figure 2: GBKN mask **Left:** Original GBKN line data; **Middle:** Final GBKN mask. **Right:** GBKN mask overlaid on LIDAR data.

respect to the surrounding field. To derive slope, a plane is fitted by least squares to a suited squared neighborhood of a LIDAR point. From the planar parameters, an estimation of the local slope is derived using Horn's method, (Burrough and McDonnell, 1998). Curvature is chosen as an attribute because the terrain at an elevation caused by a buried channel is convex as it is locally protruding. Flat terrain has a mean curvature of zero, while convex terrain has positive mean curvature. Here an approximation of mean curvature is derived from local partial derivatives by locally fitting a second degree polynomial surface to a suited squared neighborhood, see for more details (Besl and Jane, 1986) and (Nahib, 1990).

To reduce the computational costs of the least squares adjustment involved in the many slope and curvature determinations, a down-sampling strategy is applied. After an analysis of different down-sampling rates, in which slope values obtained from a down-sampled data set where compared to slope values from the full 2m grid input data set, it was decided to use only 10 % of the data.

TPI and smoothed TPI attributes The Topographic Position Index (TPI) is a measure of the elevation of a location compared to the surrounding landscape, (Weiss, 2001). To compute the TPI-value of a single pixel the difference between its elevation and the average elevation of a neighborhood around that cell is calculated. Most frequently an annular neighborhood is used, that is, all cells between a certain minimal and maximal distance are used in the calculation. A positive TPI-value means that the cell is higher than its surroundings (at the specified neighborhood size) while negative values mean it is lower. A TPI-value of zero indicates that the cell either lies on a flat area or on a constant slope. The TPI is of course strongly dependent on the scale. Here TPI-values are computed using a minimal distance of 80 m and a maximal distance of 100 m. From the TPI-values also a fourth attribute is determined, the smoothed TPI. This is just the mean of the TPI values in a 49 $\times$ 49 grid points window and helps to distinguish between small and large scale topographic features.

Maximum likelihood classification As a result of the structural attribute determination, at each grid point a 4D attribute vector is given, consisting of slope, mean curvature, TPI and smoothed TPI attribute values. The availability of these attribute vectors allows us to apply standard classification techniques from remote sensing. Here Maximum Likelihood classification is applied.

The Maximum Likelihood Classifier, (Gao, 2008), uses statistics from class signatures to determine if a given pixel belongs to a class. Each class signature is derived by manually selecting small areas that are known to belong to a certain class. These areas are called training samples. The training samples in this research have been selected based on manual interpretation of the height data and by looking at independent reference data, in this case the digital paleogeographic map, compare Fig. 1, right. In Fig. 3 the location of the training samples is shown. The results of the classification were slightly cleaned using the morphological operators 'majority filter' and 'conditional dilation' to remove small outlying classification results and fill small holes, e.g. (Jain, 1989).

3. RESULTS, VALIDATION AND DISCUSSION

In this section the results of the automatic classification of the Alblasserwaard LIDAR data are presented, validated and discussed. First visual results are discussed. Then two validation methods based on soil drillings are described, together with the results of the actual validations.

3.1 Visual validation

The blue points in Fig. 5 indicate LIDAR 2m grid points classified as 'buried channel'. Clearly some more or less connected channel structures in East-West direction are recognizable. Simultaneously, many thicker fragments classified as 'buried channel' are visible. Based on a visual evaluation it is not directly obvious if these thicker fragments indeed correspond to channel deposits. Also anomalies are visible in the classification results: thin, straight lines appear at many locations and are mainly corresponding to terrain close to roads and ditches.

Fig. 8 shows a zoom-in of the classification results, again in blue, superimposed on areal imagery data. The area in Fig. 8 approximately corresponds to the red rectangle in Fig. 5. This image

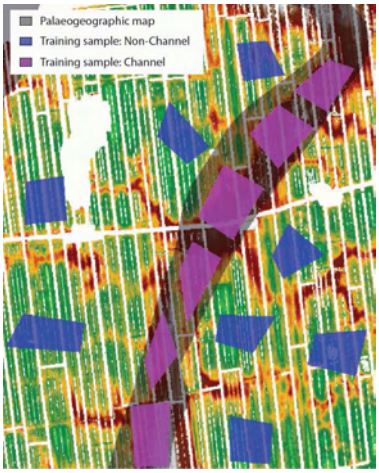


Figure 3: Training samples used for the classification process.

confirms that the classification is still influenced by infrastructure: The classification algorithm reports buried channels near and at farmyards and along a small ditch which indicates that the GBKN infrastructure database is not complete and that the filtering procedure should be further improved.

3.2 Drilling data description

For this research two independent validation data sets based on soil drillings are available. The first is a map product, the second set consist of a large amount of single drillings, interpreted by the authors.

Digital paleogeographic map A digital paleogeographic map of the complete Rhine-Meuse delta during the Holocene (including the locations of buried channel deposits) is described in (Berendsen and Stouthamer, 2001). The Alblasserwaard section of this map, Fig. 1, right, is used as validation in this research. The map is based on more than 25 years of field research using over 200 000 manual boreholes, 45 000 archaeological findings and 1 200 radiocarbon datings. The map is stored in vector format, each individual area consists of a polygon. For each area up to 12 different attributes are stored like channel size, channel length, age, year of beginning, year of ending, etc. The age of the channels on the map are given in years Before Present, where Present is defined as the year 1950. For this research four main age categories are distinguished, indicated by different colors in Fig. 1. These periods have been manually chosen based on the distribution and amount of channels abandoned in these periods.

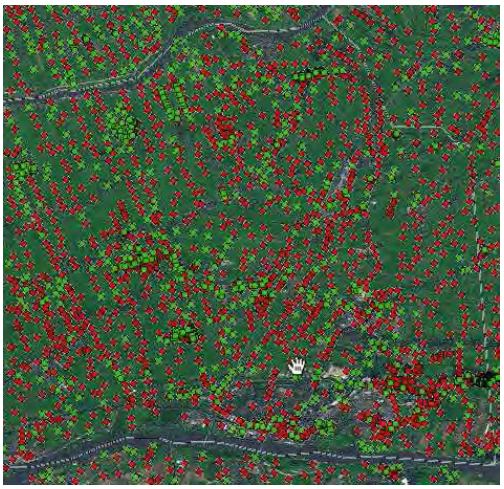


Figure 4: Classified DINO drillings.

DINO drillings The DINO database contains data and information of the subsurface of The Netherlands, (DINOLoket, 2000). The archive contains among others shallow boring measurements that are suitable to use as reference data for this research. They cover primarily the shallow subsurface and contain standardized information about the type of sediments and their depth. In total 2 680 individual drillings were available for the Alblasserwaard. The eastern part has a high drilling density, in the western part only a very limited number of drillings is available.

After importing the DINO data, each drilling was automatically analyzed to determine if the drilling was part of a buried channel deposit. This was done by applying a basic filter: search for sand layers that are cumulatively more than 3 meters in thickness in the shallow subsurface between 3 and 12 meters. If more than 3 meters of sand was found, the drilling was classified as buried channel deposits. In all other cases the drilling was classified as non channel. The reason to discard the top 3 meter is that sand layers can be present there due to other reasons, like construction works. The analysis of all of the drillings in the eastern part is shown in Fig. 4. This form of automatic interpretation of drilling data is prone to errors. This means that in this case the amount of correctly interpreted drillings is largely unknown. Still in Fig. 4 the spatial correlation between drillings and LIDAR classification results is visible.

3.3 Validation results

Further visual validation is obtained by comparing the automatically classified LIDAR points to the digital paleogeographic map and to the classified DINO drillings. For this project this was done using the ESRI Flex viewer, (ESRI, 2009). This program allows internet users to simultaneously view within their normal browser the different spatial layers on available background imagery, just as within a GIS environment. A screen shot is shown in Fig. 8. According to the digital paleogeographic map this figure contains buried channel deposits from three periods, compare Fig. 1, right. In red some relative old (6270-4621 yBP) and wide buried channels are visible, in orange another wide, slightly younger (4620-3701 yBP) channel is given. while some relatively young (3700-1700 yBP) smaller channel deposits are shown in yellow. The LIDAR data classified as channel deposit (in blue) gives the best match with the orange channel, while some matching results on the yellow channels are found as well. There seems to be hardly no correlation between the blue LIDAR channel deposits, and the large and old red channels. Similarly the classified DINO drillings give good agreement over the orange channel, while the DINO drillings give mixed responses over the red channel. Many DINO drillings outside the areas classified by the two other methods as buried channel are indeed red, but also here exceptions exist.

Table 6: Classified LIDAR vs. paleogeographic map

	Map with all channels	
LIDAR channel	3.5 %	6.3 %
non-channel	23.5 %	66.7 %
	Map 4620-1700 yBP	
LIDAR channel	2.0 %	7.6 %
non-channel	5.4 %	84.9 %
	channel	**non-channel**

These observations are partly confirmed by the numeric comparison over the region of Alblasserwaard as a whole. In Table 6, two confusion matrices of the LIDAR buried channel classification compared to the paleogeographic map are given. The top matrix compares the LIDAR classification to all buried channels shown in Fig. 1, right; in the bottom matrix the comparison is restricted to those channels in the paleogeographic map that are dated between 4620 and 1700 yBP, i.e. the yellow and orange channels in Fig. 1, right. The diagonals give the percentages of pixels where map and classified LIDAR agree, the upper right entry is the percentage of pixels that are non-channel in the reference map, but are classified as channel. The reverse holds for the bottom left entry. Although the overall classification accuracy, i.e. the trace of the confusion matrix, equals 70 %, kappa, a measure of similarity without chance agreement, only equals $\kappa = 0.06$. By removing the youngest and oldest channel class from the comparison, the amount of agreement improves to 87 % with an associated kappa value of $\kappa = 0.21$.

There are several possible reasons for this lack of agreement. First of all it should be noted that a condition for a high degree of agreement is that a buried channel deposit always results in a locally higher elevation. Although there is strong evidence that

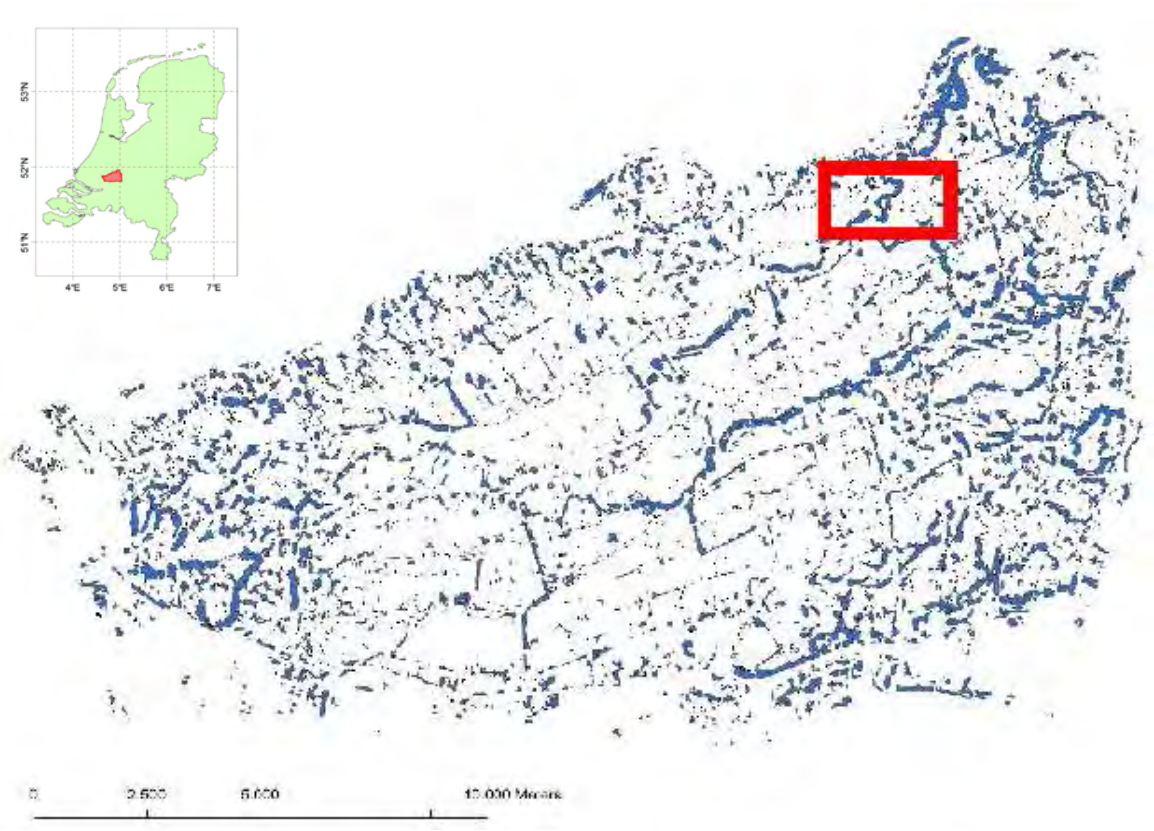


Figure 5: Buried channel classification results. The red rectangle approximately corresponds to the area of Fig. 8. The inset shows the location of the Alblasserwaard in The Netherlands.

this condition in general is fulfilled, it is not yet sufficiently clear, what buried channel characteristics result in what amount of local elevation setup. Other reasons originate in the processing of the available information. In the composition of the paleogeographic map, errors are associated to the interpretation and interpolation of the used drillings. The reason that the youngest, green, channels in the Paleogeographic map do not give a good comparison with the classified LIDAR data is simply that these channels either still exist at approximately the same location or that buildings and roads are present along or on the remains. In both cases the LIDAR data for these regions is simply filtered out in the data processing procedure.

Table 7: Classified DINO drillings vs. Classified LIDAR and paleogeographic map.

	DINO drillings	
LIDAR channel	6.3 %	7.3 %
non-channel	26.8 %	59.6 %
Map channel	15.0 %	18.6 %
non-channel	18.1 %	48.2 %
	channel	**non-channel**

In Table 7 also the confusion matrices between the classified DINO drillings and the classified LIDAR results, top, and the paleogeographic map, bottom, are given. Both the LIDAR result and the map have a comparable percentage of agreement (trace of both matrices) with the classified DINO drillings. They do however differ in the type of misclassification: in the LIDAR classification a relative large percentage of points were classified as non-channel that were channels according to our automatic interpretation of the DINO drillings. Again this could be caused by currently present infrastructure: many DINO drillings were obtained in the green zones, i.e. regions marked as young, channel deposits in the paleogeographic map, Fig. 1, right, where no reliable LIDAR surface height data is available.

4. CONCLUSIONS AND RECOMMENDATIONS

In this work, an original approach for the detection of buried channel deposits from high resolution LIDAR data has been described and validated. The first results indicate that to some extend it is possible to automatically determine the location of sand-rich channel areas: relatively young (4620-1700 yBP) and wide ($\sim$ 100m) channel deposits are often detected by the described method based on classification of a feature vector consisting of structural attributes derived from LIDAR data. Current results are however still far from a form where they could be applied in for example road construction. The results of this large case study also demonstrate that there are many assumptions/steps involved in both deriving the initial classification results and in validating these results. In future work, the impact of these assumptions on the final results should be further investigated.

Comparison to the paleogeographic map indicates that the depth, age and probably also size of the channel deposits are parameters whose influence on the relative elevation should be further investigated. The digital paleogeographic map is derived based on an interpretation of actual soil drillings and an interpolation step to connect identified channel locations to a braided network of channels. This last step has not been implemented yet for our automatic buried channel classification. To improve computational feasibility, the original LIDAR data has been downsampled in this study. To improve classification results it is recommended to start by analyzing a small area at full resolution in order to, first,

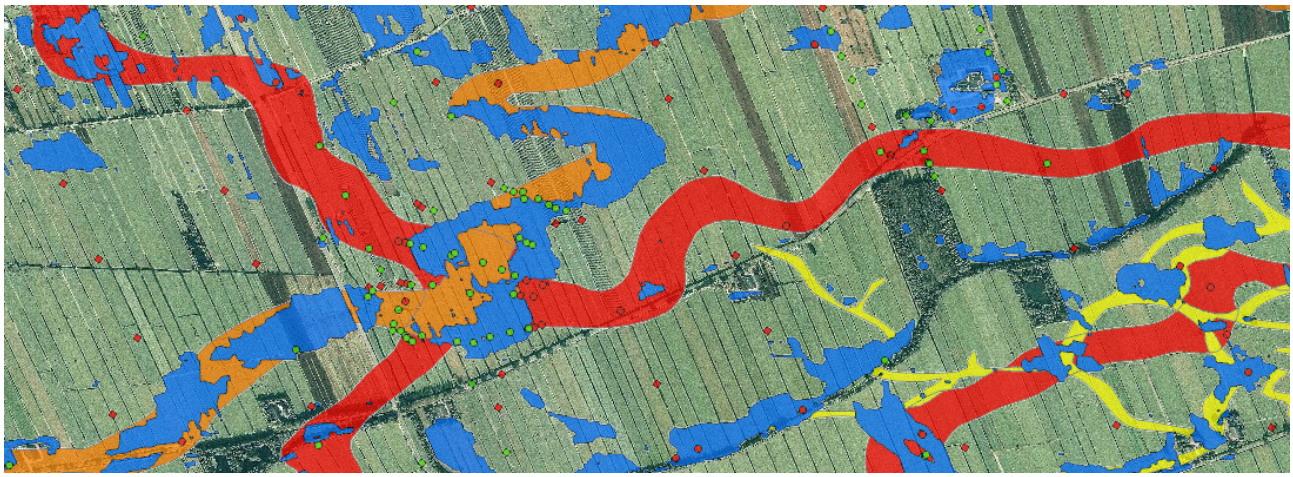


Figure 8: Buried channel classification results. **Blue:** Automatic classification LIDAR data; **Red, orange, yellow:** classification according to digital paleogeographic map, compare Fig. 1; **Red dots:** DINO drillings classified as non-channel; **Green dots:** DINO drillings classified as channels.

obtain better insight in what (channel deposit) signals are exactly present in the data and, second, to adapt the classification strategy accordingly.

REFERENCES

AHN, 2000. Productspecificaties AHN1 en AHN2. Technical report, Rijkswaterstaat, DID. http://www.ahn.nl, last visited: November 20, 2009.

Allen, J., 1965. A review of the origin and charactersitics of recent alluvial sediments. Sedimentology 5, pp. 89–101.

Berendsen, H. and Stouthamer, E., 2001. Palaeogeographic development of the Rhine-Meuse delta, The Netherlands. Koninklijke Van Gorcum.

Berendsen, H. J. A. and Volleberg, K., 2007. New prospects in geomorphological and geological mapping of the rhine-meuse delta application of detailed digital elevation maps based on laser altimetry. Netherlands Journal of Geosciences 86(1), pp. 15–22.

Besl, P. and Jane, R., 1986. Invariant surface characteristics for 3d object recognition in range images. Computer Vision, Graphics, and Image Processing archive 33(1), pp. 33–80.

Burrough, P. and McDonnell, R., 1998. Principles of Geographical Information System. Oxford University Press.

DINOLoket, 2000. Data and Information of the Subsurface of The Netherlands. Technical report, TNO. http://www.dinoloket.nl/en/DINOLoket.html, last visited: November 20, 2009.

ESRI, 2009. Get started with the sample flex viewer. Technical report, ESRI. http://www.esri.com/events/seminars/webmaps/pdfs/handout.pdf, Accessed: November 20, 2009.

Gao, J., 2008. Digital Analysis of Remotely Sensed Imagery. McGraw-Hill Professional.

GBKN, 2009. Grootschalige BasisKaart van Nederland. http://www.gbkn.nl/nieuwesite/html/engelsesite.html, Accessed: November 20, 2009.

Humme, A., Lindenbergh, R. and Sueur, C., 2006. Revealing celtic fields from lidar data using kriging based filtering. In: Proceedings ISPRS Commission V Symposium, 'Image engineering and vision metrology, Dresden.

Jain, A., 1989. Fundamentals of digital image processing. Prentice-Hall, Inc., New Jersey.

Kakiuchi, T. and Chikatsu, H., 2008. Robust extraction of ancient burial mounds in brushland from laser scanning data. IAPRS XXXVII(B5), pp. 341–346.

Kooistra, M. and Maas, G., 2008. The widespread occurrence of Celtic field systems in the central part of the Netherlands. Journal of Archaeological Science 35(8), pp. 2318–2328.

M. Doneus, C. Briese, M. F. and Jammer, M., 2008. Archaeological prospection of forested areas using full-waveform airborne laser scanning. Journal of archaeological science 35(4), pp. 882–893.

Munstermann, W., Ngan-Tillard, D. and Venmans, A., 2008. Total engineering geological approach to motorway construction on soft soils. In: Proceedings EurEnGeo conference, Madrid, Spain, pp. 1–10.

Nahib, N., 1990. Algebraic error analysis for surface curvature and segmentation of 3-d range images. Pattern Recognition 23(8), pp. 807–817.

Perron, J., Kirchner, J. and Dietrich, W., 2008. Spectral signatures of characteristic spatial scales and non-fractal structure in landscapes. Journal of Geophysical Research - Earth Surface 113(F04003), pp. 1–14.

Weiss, A., 2001. Topographic position and landforms analysis - poster presentation. In: ESRI User Conference, San Diego, United States, p. 1.

ACKNOWLEDGMENTS

The authors would like to thank Waterboard Rivierenland and Fugro Aerial Mapping B.V. for providing them with the airborne laser data. Esther Stouthamer from Utrecht University is thanked for providing the authors with a digital version of the paleogeographic map.

LAND USE AND LAND COVER CLASSIFICATION USING RADARSAT-2 POLARIMETRIC SAR IMAGE

Z. Qi [a,*], A. G. Yeh [a], X. Li [b], Z. Lin [b]

[a] Department of Urban Planning and Design, The University of Hong Kong, Pokfulam Road, Hong Kong SAR, P.R. China - qizhixin@hkusua.hku.hk
[b] School of Geography and Planning, Sun Yat-sen University, 135 West Xingang Rd., Guangzhou 510275, P.R. China - lixia@mail.sysu.edu.cn

KEY WORDS: Land cover, Land user, Polarization, RADARSAT, SAR

ABSTRACT:

Traditional pixel-based classification methods yield poor results when applied to synthetic aperture radar (SAR) imagery because of the presence of the speckle and limited spectral information in SAR data. A novel classification method, integrating polarimetric target decomposition, object-oriented image analysis, and decision tree algorithms, is proposed for land use and land cover (LULC) classification using RADARSAT-2 polarimetric SAR (PolSAR) data. The new method makes use of polarimetric information of PolSAR data, and takes advantage of object-oriented analysis and decision tree algorithms. The polarimetric target decomposition is aimed at extracting physical information from the observed scattering of microwaves by surface and volume for the classification of scattering data. The main purposes of the object-oriented image analysis are delineating objects and extracting various features, such as tone, shape, texture, area, and context. Decision tree algorithms provide an effective way to select features and create a decision tree for classification. The comparison between the proposed method and the Wishart supervised classification was made to test the performance of the proposed method. The overall accuracies of this proposed method and the Wishart supervised classification were 89.34% and 79.36%, respectively. The results show that the proposed method outperforms the Wishart supervised classification, and is an appropriate method for LULC classification of RADARSAT-2 PolSAR data.

1. INTRODUCTION

LULC classification of PolSAR data has become an important research topic since PolSAR images through ENVISAT ASAR, ALOS PALSAR and RADARSAT-2 were available. Classification methods for PolSAR images have been explored by many researchers (Rignot et al., 1992; Chen et al., 1996; Barnes and Burki, 2006; Alberga, 2007). Recently some polarimetric decomposition theorems have been introduced (Cloude and Pottier, 1996; Freeman and Durden, 1998; Yang et al., 1998; Cameron and Rais, 2006), and classification methods based on the decomposition results were also explored by some researchers (Cloude and Pottier, 1997; Lee et al., 1999; Pottier and Lee, 2000; Ferro-Famil et al., 2001). The polarimetric decomposition is aimed at extracting physical information from the observed scattering of microwaves by surface and volume for the classification of scattering data. However, so far most of the classification methods for PolSAR images are pixel-based, and have limitations for representing objects in high-resolution images and difficulties to utilize the abundant information of PolSAR imagery. The performance of pixel-based classification methods is affected by speckles, and only tonal information of pixels can be used by these methods. Moreover, the results of the pixel-based methods are almost incontinuous raster format maps, which are hardly to use to extract objects of interest and update geographical information system database expediently.

In recent years, object-oriented image analysis has been increasingly used for the classification of remote sensing data (Evans et al., 2002; Geneletti and Gorte, 2003; Li, et al., 2008). By delineating objects from remote sensing images, the object-oriented analysis can obtain a variety of additional spatial and textural information, which is important for improving the accuracy of remote sensing classification (Benz et al., 2004). However, with the addition of spatial and contextual variables, there are hundreds of features that can potentially be incorporated into the analysis. Therefore, feature selection can present a problem in object-based classification (Laliberte et al., 2006). Since recently some polarimetric decomposition theorems have been introduced, which brings abundant polarimetric information, the problems of feature selection become more intractable.

Decision tree algorithms can be used to solve the problems of feature selection. Decision trees are commonly used for variable selection to reduce data dimensionality in image analysis (Lawrence and Wright, 2001). Classification accuracies from decision tree classifiers are often greater compared to using maximum likelihood or linear discriminant function classifiers (Laliberte et al., 2006). Some studies have indicates that decision trees can provide an accurate and efficient methodology for classification of remote sensing data (Friedl and Brodley, 1997; Swain and Hauska, 1977; Mclver and Friedl, 2002).

The objective of this study is to explore a new classification method of integrating polarimetric decomposition, object-oriented image analysis, and decision tree algorithms for PolSAR imagery. In this method, 39 polarimetric descriptors were first calculated by using the H/A/Alpha polarimetric decomposition and combined with the parameters of the scattering and coherency matrix to form a multichannel image. Next, during the object-oriented image analysis, image objects were delineated by implementing a multi-resolution segmentation on the Pauli composition image of RADARSAT-2 PolSAR data. Meanwhile, a total of 1253 features were extracted for each image object. Then, a decision tree algorithm was used to select features and create a decision tree for the classification. Finally, the LULC classification of RADARSAT-

* Corresponding author

2 PolSAR image was carried out by using the constructed decision tree.

2. METHODOLOGY

2.1 Polarimetric target decomposition

Target decomposition theorems were first formalized by Huynen (Huynen, 1970), and then many decomposition methods were proposed by other researchers. In this study we just focus on H/A/Alpha decomposition (Cloude and Pottier, 1997). The H/A/Alpha decomposition is also called eigenvector-eigenvalue based decomposition. This decomposition method is free from the physical constraints imposed by multivariate models because it is not dependent on the assumption of a particular underlying statistical distribution.

The most important observable measured by radar system is the 3×3 coherency matrix $\boldsymbol{T_3}$, and it is written as:

$$T_3 = \frac{1}{2}\begin{bmatrix} |S_{hh}+S_{vv}|^2 & (S_{hh}+S_{vv})(S_{hh}-S_{vv})^* & 2(S_{hh}+S_{vv})S_{hv}^* \\ (S_{hh}-S_{vv})(S_{hh}+S_{vv})^* & |S_{hh}-S_{vv}|^2 & 2(S_{hh}-S_{vv})S_{hv}^* \\ 2S_{hv}(S_{hh}+S_{vv})^* & 2S_{hv}(S_{hh}-S_{vv})^* & 4|S_{hv}|^2 \end{bmatrix} \quad (1)$$

where, S_{hh} and S_{vv} denote the copolarized complex scattering amplitudes; S_{hv} and S_{vh} denote the cross-polarized complex scattering amplitudes, respectively; * denotes conjugate, and | | denotes module.

According to the H/A/Alpha decomposition theorem, $\boldsymbol{T_3}$ can be decomposed as follows:

$$T_3 = \begin{bmatrix} \underline{u}_1 & \underline{u}_2 & \underline{u}_3 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} \underline{u}_1 & \underline{u}_2 & \underline{u}_3 \end{bmatrix}^{-1} \quad (2)$$

where $\lambda_1\ \lambda_2\ \lambda_3$ are the eigenvalues of $\boldsymbol{T_3}$, and $\infty>\lambda_1>\lambda_2>\lambda_3>0$. $\underline{u}_i$ for i=1,2,3 are the eigenvectors of $\boldsymbol{T_3}$ and can be formulated as follows:

$$\underline{u}_i = \begin{bmatrix} \cos\alpha_i & \sin\alpha_i \cos\beta_i e^{j\delta_i} & \sin\alpha_i \cos\beta_i e^{jr_i} \end{bmatrix}^T \quad (3)$$

where, the symbol T stands for complex conjugate.

The pseudo-probabilities of the $\boldsymbol{T_3}$ expansion elements are defined from the set of sorted eigenvalues.

$$p_i = \lambda_i \Big/ \sum_{k=1}^{3} \lambda_k \quad \text{with} \quad p_1 \geq p_2 \geq p_3 \quad (4)$$

where p_i is called the probability of the eigenvalue λ_i and represents the relative importance of this eigenvalue to the total scattered power.

In general, the columns of the 3×3 unitary matrix $\boldsymbol{U_3}$ are not only unitary but mutually orthogonal. This means that in practice, the parameters (α_1, α_2, α_3), (β_1, β_2, β_3), (δ_1, δ_2, δ_3), and (γ_1, γ_2, γ_3) are not independent. Thus, each polarimetric parameter is associated with a three-symbol Bernoulli statistical process. In this way, the estimate of the mean polarimetric parameter set is given by (Lee and Poitter, 2009):

$$(\overline{\alpha}, \overline{\beta}, \overline{\gamma}, \overline{\delta}) = \sum_{i=1}^{3} p_i \cdot (\alpha_i, \beta_i, \gamma_i, \delta_i) \quad (5)$$

The eigenvalues and eigenvectors are the primary parameters of the eigenvector-eigenvalue based decomposition of $\boldsymbol{T_3}$. The different secondary polarimetric descriptors are defined as a function of the eigenvalue and eigenvetors of $\boldsymbol{T_3}$ for simplifying the analysis of the physical information provided by this decomposition (Lee and Poitter, 2009):

1) Polarimetric scattering entropy (H)
The polarimetric scattering entropy defines the degree of statistical disorder of each distinct scatter type within the ensemble.

2) Polarimetric scattering anisotropy (A)
The anisotropy measures the relative importance of the second and the third eigenvalues of the eigen decomposition. It is a useful parameter to improve the capability to distinguish different types of scattering process.

3) Combinations between entropy (H) and anisotropy (A)
- The (1- H)(1- A) image corresponds to the presence of a single dominant scattering process.
- The H(1- A) image characterizes a random scattering process.
- The HA image relates to the presence of two scattering mechanisms with the same probability.
- The (*1- H*)*A* image corresponds to the presence of two scattering mechanisms with a dominant process and a second one with medium probability.

4) SERD and DERD
Single bounce Eigenvalue Relative Difference (SERD) and the Double bounce Eigenvalue Relative Difference (DERD) are sensitive to natural media characteristics and can be used for quantitative inversion of bio- and geophysical parameters.

5) Polarization asymmetry (PA) and Polarization fraction (PF)
PF ranges between 0 and 1, when λ_3=0 the entire return is polarized, and when λ_3>0 the polarization fraction drops. The PA measures the relative strength of the two polarimetric scattering mechanisms.

6) Radar vegetation Index (RVI) and Pedestal height (PH)
The RVI was used for analyzing scattering from vegetated areas. The PH is another polarization signature of measuring randomness in the scattering.

7) Target randomness parameter (P_R)
P_R is very close to the entropy (H) and provides the similar information.

8) Shannon entropy (SE):
This parameter is a sum of SE_I and SE_P. SE_I is the intensity contribution that depends on the total backscattered power, and SE_P the polarimetric contribution that depends on the Barakat degree of polarization.

The information provided by the H/A/Alpha decomposition of the coherency matrix can be interrelated in terms of the eigenvalues and eigenvectors of the decomposition or in terms of secondary polarimetric descriptors. Both interpretations have to be considered as complementary. In this study, PolSARPro 4.0 software was used to implement the H/A/Alpha decomposition (López-Martínez, 2005).

2.2 Object-oriented image analysis

At present most of orbital radar images are obtained using only one single frequency. There are significant confusions if classification is purely based on a single band of SAR images (Li and Yeh, 2004). One way to compensate for the limited information of single frequency is to derive more features such as texture and shape for the classification beside of spectral information of radar images. Object-oriented image analysis can be used on radar images to extract such type of information. Such analysis is especially useful in improving the classification accuracy of high-resolution images (Langford, 2002).

There are two steps in the object-oriented analysis: (1) Image segmentation; (2) Feature extraction. In this study, multi-resolution segmentation was used to delineate objects based on shape and color homogeneity (Benz et al., 2004) from the Pauli composition image of RADARSAT-2 PolSAR data. Pauli composition has become today the standard for PolSAR image display, and thus it is often used for visual interpretation and observation. In order to obtain the best discrimination results, both color and shape heterogeneity is used to adjust the segmentation. The object-oriented package, Definiens Developer 7.0 (previously called eCognition), is used to implement the object-oriented image analysis.

The multi-resolution segmentation algorithm is a bottom up region-merging technique starting with one-pixel objects. During the region-merging process, smaller image objects are merged into bigger ones, and the underlying optimization procedure minimizes the weighted heterogeneity of resulting image objects for a given resolution over the whole scene. Heterogeneity is determined based on the standard deviation of color properties and their shapes. The merging of a pair of adjacent image objects leads to the smallest growth of the defined heterogeneity. The process will stop if the smallest growth exceeds the threshold defined by a scale parameter.

Various types of features can be obtained via object-oriented image analysis, such as mean value of pixels and standard deviation of pixels in an image object, mean difference to neighbors, area, length, GLCM Homogeneity, GLCM Contrast and so on. The interrelations among objects can be defined and utilized as additional information for classification.

2.3 Decision tree algorithm

The task of this work is to determine the class of each image object based on their features. Since a large set of features can be extracted from image objects, the selection of proper features is important for classification. Decision trees are commonly used for variable selection to reduce data dimensionality in image analysis (Lawrence and Wright, 2001). Decision trees are used to predict membership of cases or objects in the classes of a categorical dependent variable from their measurements on one or more predictor variables. In these tree structures, leaves represent classifications and branches represent conjunctions of features that lead to those classifications.

Decision tree algorithms have many advantages: (1) they are white box model and simple to understand and interpret. If a given result is provided by the model, the explanation for the result is easily replicated by simple math; (2) Decision trees are able to handle both numerical and categorical data, and requires little data preparation; (3) They are robust and perform well with large data in a short time; (4) Decision trees, performing univariate splits and examining the effects of predictors one at a time, have implications for the variety of types of predictors that can be analyzed.

In this study, QUEST was used to implement the LULC classification. The QUEST is a binary-split decision tree algorithm for classification and data mining (Loh and Shih, 1997). A decision tree can be created based on training samples using QUEST. After the decision tree is constructed, it can be used to identify the class of other unknown cases.

3. IMPLEMENTATIONS AND RESULTS

3.1 The study area and data processing

The study area is located in the Panyu District with latitudes 22º51´ to 22º58´ and longitudes 113º20´ to 113º33´ of Guangzhou in southern China. Panyu lies at the heart of the Pearl River Delta, and has a total land area of 1,314 km^2 and a population of 926,542. This district was an agricultural country before economic reform in 1978 but has been transformed recently into a rapidly urbanized area. Since Panyu became a district of Guangzhou in July 2000, intensive land development has occurred to provide housing to the residents of Guangzhou City. Huge profits have been generated through property development, which results in the increase of land speculation activities and illegal land development. Accurate and timely LULC information can provide government with scientific information for making management policies to control and prevent illegal developments at an early stage.

The RADARSAT-2 image with the fine quad-pol (FQ12) and Single Look Complex (SCL) obtained on 21 March 2009 was used in this study (Figure 1.). The image has a full polarization of HH, HV, VH and VV, a resolution of 5.2 × 7.6m and an incidence angle of 31.5 degrees. Data processing included radiometric calibration, geometric calibration, slant range to ground range and image filtering. Lee refined filter, which proves to be efficient in polarimetric SAR images, was applied to the RADARSAT-2 data.

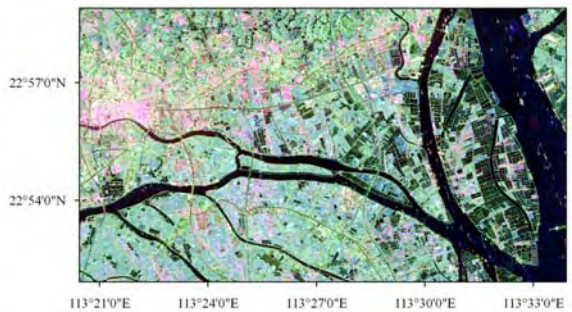


Figure 1. RADARSAT-2 Quad-Pol image of the study area (Pauli Composition: HH+VV, HV, HH-VV)

The final LULC map displays categories including built-up area, water, barren land and vegetation. In field work, a total of 497 field plots were selected across typical LULC classes based on a

clustered sampling approach (McCoy, 2005). A suggestion by Congalton and Green (1999), based on experience with the multinomial distribution, is to use a minimum of 50 samples for each category. A minimum sample unit should be no smaller than 3×3 cluster of pixels or a polygon of comparable size for either training sites or accuracy assessment sites (McCoy, 2005). In this study, the sampling size per field plot ranged from 15 to 54 pixels, which was determined by the size of a ground photos taken concurrently. The collected field plots were divided into two groups for the training and validation. There were 210 plots in the training group and 287 plots in the validation group. The first group was used to select features and create a decision tree with QUEST for the classification, and the second group, which included 7520 pixels, was used to verify the results of the LULC classification. An ALOS image of the 10m multispectral bands obtained on 31 November 2008 was used as a reference map to assist the collection of ground true information.

3.2 Polarimetric target decomposition results

The H/A/Alpha decomposition was used to extract polarimetric descriptors from the RADARSAT-2 data. A total of 39 descriptors were extracted and combined with the elements of the scattering and coherency matrix to form a PolSAR multichannel image. The descriptors and the corresponding image channels are listed in Table 2.

Channel	Descriptor	Channel	Descriptor	Channel	Descriptor
1	HH	19	p_2	37	*H*
2	HV	20	p_3	38	*A*
3	VH	21	$\overline{\alpha}$	39	(1-*H*)(1-*A*)
4	VV	22	α_1	40	(1-*H*)*A*
5	HH+VV	23	α_2	41	*H*(1-*A*)
6	HV+VH	24	α_3	42	*HA*
7	HH-VV	25	$\overline{\beta}$	43	SERD
8	T11	26	β_1	44	DERD
9	T12	27	β_2	45	PA
10	T13	28	β_3	46	PF
11	T22	29	$\overline{\delta}$	47	RVI
12	T23	30	δ_1	48	PH
13	T33	31	δ_2	49	P_R
14	$\overline{\lambda}$	32	δ_3	50	SE
15	λ_1	33	$\overline{\gamma}$	51	SE_I
16	λ_2	34	γ_1	52	SE_P
17	λ_3	35	γ_2		
18	p_1	36	γ_3		

Table 2. Descriptors and corresponding channels of the PolSAR multichannel image

3.3 Image segmentation results

The multi-resolution segmentation was used to delineate image objects and extract their features. The scale parameter determines the maximum change in heterogeneity that may occur when merging two image objects. Adjusting of the value of scale parameter influences the average object size. A larger value leads to bigger objects and vice versa. The optimal scale parameters for the multi-resolution segmentation were found according to some experiments. The corresponding segmentation results related to different segmentation scale parameters are shown in Figure 3. The segmentation with a scale of 10 was enough to delineate accurate land parcels.

Since the combined image consists of 52 channels, the number of features that can be extracted from one single object is as

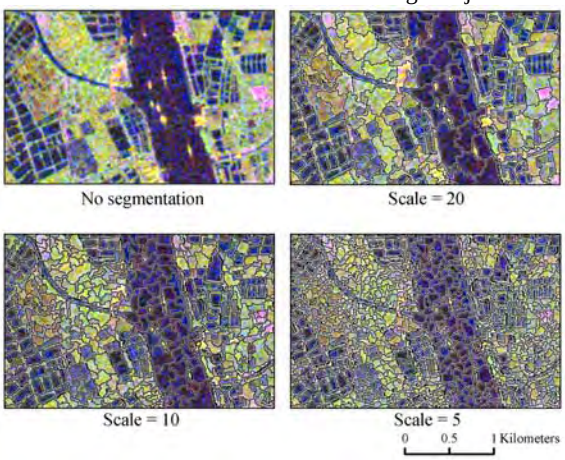


Figure 3. Determining the optimal scale for segmentation of Radarsat-2 image.

high as 1253. These features are the indigenous parameters of Definiens Developer 7.0, and they are listed as the following four major categories:

1) 208 (4×52) indicators related to the statistical values of each object: min, max, mean, and standard deviation of each layer;
2) 624 (12×52) indicators related to texture (e.g., Grey-level co-occurrence matrix (GLCM) Homogeneity, GLCM Contrast, GLCM Dissimilarity, and GLCM Entropy);
3) 364 (7×52) indicators related to spatial relationship (e.g., mean difference to neighbors, and mean difference to brighter neighbors);
4) 57 indicators related to shape (e.g., area, length, number of segments, and curvature/length (only main line) were extracted from an object;

3.4 Land use and land cover classification results

The classification maps using the proposed method and the Wishart supervised classification were produced as the comparison (Figure 4). The Wishart supervised classification was implemented by using the PolSARPro 4.0 software (López-Martínez, 2005).

3.5 Accuracy evaluation

Accuracy evaluation was carried out based on field investigation. The accuracy statistics of these two methods were produced in Table 5 and 6. The overall accuracy of our method was 89.34%, which was much higher than the Wishart supervised classification method, with an overall accuracy of 79.36%. Moreover, the overall kappa accuracy was also increased from 72.41% to 85.76% by using this proposed approach. The commission and omission errors can be measured by the user's and procurer's accuracy. The proposed method distinguished each class better than the Wishart supervised classification did, especially for identifying the built-up area. In the Wishart supervised classification, the accuracy of the user for the built-up area is very poor (50.28%). In this proposed method, however, the user's accuracy of the built-up area was much higher (77.84%). In the classification, industrial buildings with wide flat roof were commonly confused with barren land because of similar scattering mechanism. Some shadow of buildings was also prone to be classified as barren land. Otherwise, some buildings that have specific orientations not aligned in the azimuth direction or have complex structures

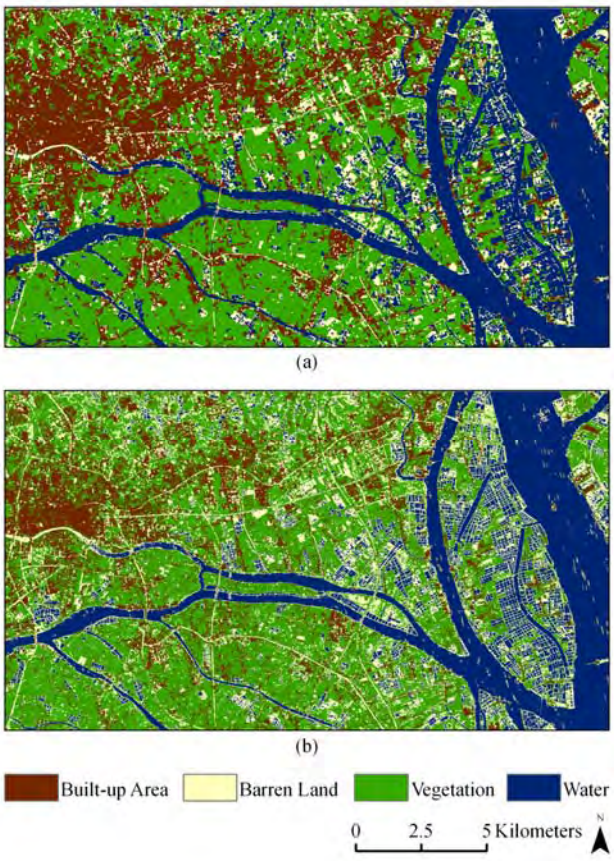


Figure 4. Land use and land cover classification results (a) Proposed method, (b) Wishart supervised classification.

such as rough roofs were assimilated into volume scattering class and then assigned to the vegetation class. In some mountain area covered with vegetation, some shadow was also prone to be classified as barren land. Barren lands with water on their surface or high soil moisture were misclassified as water in the classification result. Some polluted water areas were also assigned to the barren land.

4. CONCLUSIONS

This paper proposed a new method that integrates polarimetric decomposition, object-oriented image analysis, and decision tree algorithms. The comparison between the proposed method and the Wishart supervised classification method indicates that the proposed method outperforms the Wishart supervised classification method, and can reduces incontinuous phenomenon effectively. The results show that the overall classification accuracy of the proposed method was 89.34% whereas it was 79.36% in the Wishart supervised classification method. Moreover, the overall kappa accuracy of the proposed method was also higher than that of the Wishart supervised classification method. The object-oriented image analysis is the suitable image processing method for information extraction to support the classification of polarimetric SAR images. Decision tree algorithms are efficient tools for the object-oriented classification of polarimetric SAR image. The experiments have indicated that the proposed method is an appropriate method for LULC classification of polarimetric SAR imagery.

REFERENCES

Alberga, V., 2007. A study of land cover classification using polarimetric SAR parameters. *International Journal of Remote Sensing*, 28(17), pp. 3851-3870.

Barnes, C. F. and Burki, J., 2006. Late-season rural land-cover estimation with polarimetric-SAR intensity pixel blocks and sigma-tree-structured near-neighbor classifiers. *IEEE Transactions on Geoscience and Remote Sensing*, 44(9), pp. 2384-2392.

	Reference data					User's
Classification data	Water	Vegetation	Built-up Area	Barren Land	Total	Accuracy (%)
Water	1,888	0	0	32	1,920	98.33
Vegetation	0	1,838	66	16	1,920	95.73
Built-up Area	19	234	1,370	137	1,760	77.84
Barren Land	173	86	39	1,622	1,920	84.48
Total	2,080	2,158	1,475	1,807	7,520	
Producer's accuracy (%)	90.77	85.17	92.88	89.76		
Overall accuracy (%)	89.34					
Overall kappa accuracy (%)	85.76					

Table 5. Classification accuracy (%) of the proposed method

	Reference data					User's
Classification data	Water	Vegetation	Built-up Area	Barren Land	Total	Accuracy (%)
Water	1,877	0	0	43	1,920	97.76
Vegetation	0	1,590	250	80	1,920	82.81
Built-up Area	2	666	885	207	1,760	50.28
Barren Land	221	44	39	1,616	1,920	84.17
Total	2,100	2,300	1,174	1,946	7,520	
Producer's accuracy (%)	89.38	69.13	75.38	83.04		
Overall accuracy (%)	79.36					
Overall kappa accuracy (%)	72.41					

Table 6. Classification accuracy (%) of the Wishart supervised classification

Benz, U. C., Hofmann, P., Willhauck, G., Lingenfelder, I. and Heynen, M., 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58, pp. 239-258.

Cameron, W. L. and Rais, H., 2006. Conservative polarimetric scatterers and their role in incorrect extensions of the Cameron decomposition. *IEEE Transactions on Geoscience and Remote Sensing*, 44(12), pp. 3506-3516.

Chen, K. S., Huang, W. P., Tsay, D. H. and Amar, F., 1996. Classification of multifrequency polarimetric SAR imagery using a dynamic learning neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 34(3), pp. 814-820.

Cloude, S. R. and Pottier, E., 1996. A review of target decomposition theorems in radar polarimetry. *IEEE Transactions on Geoscience and Remote Sensing*, 34(2), pp. 498-518.

Cloude, S. R. and Pottier, E., 1997. An entropy based classification scheme for land applications of polarimetric SAR. *IEEE Transactions on Geoscience and Remote Sensing*, 35(1), pp. 68-78.

Congalton, R. and Green, K., 1999. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. CRC Press, Boca Raton, pp. 74-79.

Evans, C., Jones, R., Svalbe, I. and Berman, M., 2002. Segmenting multispectral Landsat TM images into field units. *IEEE Transactions on Geoscience and Remote Sensing*, 40(5), pp. 1054-1064.

Ferro-Famil, L., Pottier, E. and Lee, J. S., 2001. Unsupervised classification of multifrequency and fully polarimetric SAR images based on the H/A/Alpha-Wishart classifier. *IEEE Transactions on Geoscience and Remote Sensing*, 39(11), pp. 2332-2342.

Freeman, A. and Durden, S. L., 1998. A three-component scattering model for polarimetric SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 36(3), pp. 963-973.

Friedl, M. A. and Brodley, C. E., 1997. Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, 61(3), pp. 399-409.

Geneletti, D. and Gorte, B. G. H., 2003. A method for object-oriented land cover classification combining Landsat TM data and aerial photographs. *International Journal of Remote Sensing*, 24(6), pp. 1273-1286.

Huynen, J. R., 1970. Phenomenological theory of radar targets. *PhD dissertation*. Drukkerij Bronder-offset N. V., Rotterdam.

Laliberte, A. S., Koppa, J., Fredrickson, E. L. and Rango, A., 2006. Comparison of nearest neighbor and rule-based decision tree classification in an object-oriented environment. *2006 IEEE International Geoscience and Remote Sensing Symposium*, Vols 1-8, 3923-3926.

Langford, W. T., 2002. *Evaluation Functions in Aerial Image Segmentation*, Umi Proquest Digital Dissertations, Oregon State University, pp. 50-84.

Lawrence, R. L. and Wright, A., 2001. Rule-based classification systems using classification and regression tree (CART) analysis. *Photogrammetric Engineering and Remote Sensing*, 67(10), pp. 1137-1142.

Lee, J. S., Grunes, M. R., Ainsworth, T. L., Du, L. J., Schuler, D. L. and Cloude, S. R., 1999. Unsupervised classification using polarimetric decomposition and the complex Wishart classifier. *IEEE Transactions on Geoscience and Remote Sensing*, 37(5), pp. 2249-2258.

Lee, J. S. and Pottier, E., 2009. Polarimetric Radar Imaging from Basics to Applications. (p. 180). CRC Press, New York.

Li, X. and Yeh, A. G., 2004. Multi-temporal SAR images for monitoring cultivation systems using case-based reasoning. *Remote Sensing of Environment*, 90(4), 524-534.

Li, H. T., Gu, H. Y., Han, Y. S. and Yang, J. H., 2008. Object-oriented classification of polarimetric SAR imagery based on statistical region merging and support vector machine. *2008 International Workshop on Earth Observation and Remote Sensing Applications*, pp. 147-152.

Loh, W. Y. and Shih, Y. S., 1997. Split selection methods for classification trees. *Statistica Sinica*, 7(4), pp. 815-840.

López-Martínez, C., Ferro-Famil, L., and Pottier, E., 2005. PolSARpro v4.0 Polarimetry Tutorial. http://earth.esa.int/polsarpro/tutorial.html.

Mccoy, R.M., 2005, *Field Methods in Remote Sensing*. The Guilford Press, New York.

Mclver, D. K. and Friedl, M. A., 2002. Using prior probabilities in decision-tree classification of remotely sensed data. *Remote Sensing of Environment*, 81, pp. 253-261.

Pottier, E. and Lee, J. S., 2000. Application of the "H/A/Alpha" polarimetric decomposition theorem for unsupervised classification of fully polarimetric SAR data based on the Wishart distribution. *CEOS SAR Workshop*, Toulouse, France, 450, pp. 335-340.

Rignot, E., Chellappa, R. and Dubois, P., 1992. Unsupervised segmentation of polarimetric SAR data using the covariance-matrix. *IEEE Transactions on Geoscience and Remote Sensing*, 30(4), pp. 697-705.

Swain, P. H. and Hauska, H., 1977. Decision tree classifier: design and potential. *IEEE Transactions on Geoscience and Remote Sensing*, 15(3), pp. 142-147.

Yang, J., Yamaguchi, Y., Yamada, H., Sengoku, M. and Lin, S. M., 1998. Stable decomposition of mueller matrix. *IEICE Transactions on Communications*, E81-B(6), pp. 1261-1268.

ACKNOWLEDGEMENT

This work was supported by the Science and Operational Applications Research for RADARSAT-2 Program (SOAR 2762). The authors would like to thank MDA GEOSPATIAL SERVICES INC., Canada for providing the RADARSAT-2 data.

FOUR REDUCED-REFERENCE METRICS FOR MEASURING HYPERSPECTRAL IMAGES AFTER SPATIAL RESOLUTION ENHANCEMENT

S. Qian [a,] *, G. Chen [a]

[a] Canadian Space Agency, 6767 route de l'Aeroport, St-Hubert, Quebec, Canada J3Y8Y9
(shen-en.qian, guangyi.chen)@asc-csa.gc.ca

KEY WORDS: Hyper spectral, Image, Spatial, Resolution, Metric

ABSTRACT:

In this paper, four new reduced-references (RR) metrics are proposed for measuring the visual quality of hyperspectral images after having undergone spatial resolution enhancement. These metrics can measure the visual quality of hyperspectral images whose full-reference (FR) image is not available whereas the low spatial resolution reference image is available. A FR metric requires the reference image and the test image to have the same size. After spatial resolution enhancement of hyperspectral images, the size of the enhanced images is larger than that of the original image. Thus, the FR metric cannot be used. A common approach in practice is to first down-sample an original image to a low resolution image, then to spatially enhance the down-sampled low resolution image using an enhancement technique. In this way, the original image and the enhanced image have the same size and the FR metric can be applied to them. However, this common approach can never directly assess the image quality of the spatially enhanced image that is produced directly from the original image. Experimental results showed that the proposed RR metrics work well for measuring the visual quality of spatial resolution enhanced hyperspectral images. They are consistent with the corresponding FR metrics.

1. INTRODUCTION

Measurement of image quality is of fundamental importance to many image processing applications. Image quality assessment algorithms are in general classified into three categories: full-reference (FR), reduced-reference (RR), and no-reference (NR) algorithms. True NR algorithms are very difficult to design and little progress has been made (Sheikh et al, 2005). FR algorithms are easier to design and the majority of image quality assessment algorithms are of this type. In FR quality assessment, a reference image of perfect quality is assumed to be available. However, in RR or NR quality assessment, partial or no reference information is available.

Mean square error (MSE) is the simplest FR metric between the reference image x and the processed image y:

$$MSE(x,y)=\frac{1}{N}\sum_{i,j}(y(i,j)-x(i,j))^2 \qquad (1)$$

where N is the total number of pixels in the images x and y. The MSE is easy to compute and implement in software and hardware. However, the MSE is not a good image quality measure as it is not well matched to perceived image quality. Two distorted images with the same MSE may have very different types of errors, some of which are more visible than others. Thus one image may look very much similar to the reference, whereas another may look very much distorted.

Peak signal to noise ratio (PSNR) is also a popular FR metric to measure the quality of a reconstructed image, and it is defined as:

$$PSNR(x,y)=10\log_{10}\left(\frac{\max(x)^2}{\frac{1}{N}\sum_{i,j}(y(i,j)-x(i,j))^2}\right) \qquad (2)$$

The PSNR has been used as a standard metric in image denoising and other related image processing tasks.

Wang and Bovik (2002) proposed the Q index for a reference image x and an image y to be evaluated,

$$Q(x,y)=\frac{\sigma_{xy}}{\sigma_x\sigma_y}\cdot\frac{2\mu_x\mu_y}{\mu_x^2+\mu_y^2}\cdot\frac{2\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2} \qquad (3)$$

where μ_x and μ_y are sample means, σ_x^2 and σ_y^2 are sample variances, and σ_{xy} is the sample cross-covariance between x and y. The Q index is a FR metric and it is easy to calculate and applicable to various image processing applications. It outperforms the MSE significantly under different types of image distortions. Wang et al (2004) also developed the structural similarity (SSIM) index, which is also a FR metric, by comparing local correlations in luminance, contrast, and structure between the reference and distorted images. The SSIM index is defined as:

$$SSIM(x,y)=\frac{\sigma_{xy}+C_1}{\sigma_x\sigma_y+C_1}\cdot\frac{2\mu_x\mu_y+C_2}{\mu_x^2+\mu_y^2+C_2}\cdot\frac{2\sigma_x\sigma_y+C_3}{\sigma_x^2+\sigma_y^2+C_3} \qquad (4)$$

where μ_x and μ_y are sample means of images x and y, σ_x^2 and σ_x^2 are sample variances, and σ_{xy} is the sample cross-covariance between x and y. The constants C_1, C_2, C_3 stabilize SSIM when the means and variances become small. The mean SSIM (MSSIM) over the whole image gives the final quality measure.

* Corresponding author

Sheikh and Bovik (2006) developed a visual information fidelity (VIF) index for FR measurement of image visual quality. Let $\boldsymbol{e}=\boldsymbol{c}+\boldsymbol{n}$ be the reference image, and $\boldsymbol{n}$ zero-mean normal distribution $N(0,\sigma_n^2 I)$ noise. Also, let $\boldsymbol{f}=\boldsymbol{d}+\boldsymbol{n}'=\boldsymbol{g}\boldsymbol{c}+\boldsymbol{v}'+\boldsymbol{n}'$ be the test image, where $\boldsymbol{g}$ represents the blur, $\boldsymbol{v}'$ the additive zero-mean Gaussian white noise with covariance $\sigma_v^2 I$, and $\boldsymbol{n}'$ the zero-mean normal distribution $N(0,\sigma_n^2 I)$ noise. Then, VIF can be computed as the ratio of the mutual information between $\boldsymbol{c}$ and $\boldsymbol{f}$, and the mutual information between $\boldsymbol{c}$ and $\boldsymbol{e}$ for all wavelet sub-bands except the lowest approximation subband.

$$VIF = \frac{\sum I(c; f \mid z)}{\sum I(c; e \mid z)} \tag{5}$$

All the metrics above are popular metrics published in the literature for FR image quality assessment. However, they require the reference image and the test image to have the same image size. After spatial resolution enhancement of hyperspectral images, the size of the enhanced images is larger than that of the original image. Thus, these metrics cannot be used to assess the quality of the enhanced images. A common approach in practice is to first down-sample an original image to a low resolution image, then to spatially enhance the down-sampled low resolution image using an enhancement technique. In this way, the original image and the enhanced image have the same size and the FR metrics can be applied to them. However, this common approach can never directly assess the image quality of the spatially enhanced image that is produced directly from the original image. The image quality of the enhanced image measured based on the down-sampled low resolution image may or may not reflect the real quality of the image that is enhanced directly from the original image, as the down-sampling procedure introduces artificial effects.

This paper proposes new RR metrics. A brief review about the RR metric is given here. Wang and Simoncelli (2005) proposed an RR image quality assessment method based on a natural image statistic model in the wavelet transform domain. They used the Kullback-Leibler distance between the marginal probability distributions of wavelet coefficients of the reference and distorted images as a measure of image distortion. A generalized Gaussian model was employed to summarize the marginal distribution of wavelet coefficients of the reference image, so that only a relatively small number of RR features are needed for the evaluation of image quality. Li and Wang (2009) proposed an RR algorithm using statistical features extracted from a divisive normalization-based image representation. They demonstrated that such an image representation has simultaneous perceptual and statistical relevance and its statistical properties are significantly changed under different types of image distortions. Engelkea et al (2009) developed RR objective perceptual image quality metrics for use in wireless imaging. Instead of focusing only on artifacts due to source encoding, they followed an end-to-end quality approach that accounts for the complex nature of artifacts that may be induced by a wireless communication system.

In this paper, four new RR metrics were proposed for measuring the image fidelity of a testing image that has higher spatial resolution (i.e. larger size than that of the original image). It is assumed that a low spatial resolution reference image is available, whereas the high spatial resolution reference image is not. These four proposed RR metrics do not require the sizes of the reference image and the test image to be the same.

The iterative back projection (IBP) (Irani and Peleg, 1991, 1993) technique was chosen to enhance the spatial resolution of testing hyperspectral images in order to demonstrate the usefulness of these metrics. Experimental results reported in section 3 show that the proposed metrics can measure the image quality of the spatial resolution enhanced images very well.

2. CONSTRUCTING NEW RR METRICS FROM EXISTING FR METRICS

In this section, four new RR metrics are proposed for assessing the image quality of a spatial resolution enhanced image. They can be derived as follows.

Let the size of the low spatial resolution image $\boldsymbol{f}$ be $P\times Q$, and the size of the corresponding spatial resolution enhanced image $\boldsymbol{g}$ be $2P\times 2Q$. This means that the spatial resolution of image $\boldsymbol{f}$ is enhanced at a factor of 2×2. The following four down-sampled images at a factor of 2×2, can be defined as:

$$g_{11} = g(1:2:2P, 1:2:2Q) \tag{6}$$
$$g_{12} = g(1:2:2P, 2:2:2Q) \tag{7}$$
$$g_{21} = g(2:2:2P, 1:2:2Q) \tag{8}$$
$$g_{22} = g(2:2:2P, 2:2:2Q) \tag{9}$$

where $g(i:2:2P,\ j:2:2Q)$, ($i=1,2;\ j=1,2$), is a matrix which starts at the pixel (i,j) of image $\boldsymbol{g}$ and extract every other pixels in $\boldsymbol{g}$ along both the x and the y directions with a step of 2. Since the low spatial resolution image $\boldsymbol{f}$ and the images $g_{i,j}$ ($i,j=1,2$) have the same image size, one can use any FR metrics to measure the image quality between them. The following four RR metrics are proposed in this paper:

$$PSNR(f;g) = \frac{1}{4}\sum_{i=1}^{2}\sum_{j=1}^{2} PSNR(f;g_{ij}) \tag{10}$$

$$Q(f;g) = \frac{1}{4}\sum_{i=1}^{2}\sum_{j=1}^{2} Q(f;g_{ij}) \tag{11}$$

$$MSSIM(f;g) = \frac{1}{4}\sum_{i=1}^{2}\sum_{j=1}^{2} MSSIM(f;g_{ij}) \tag{12}$$

$$VIF(f;g) = \frac{1}{4}\sum_{i=1}^{2}\sum_{j=1}^{2} VIF(f;g_{ij}) \tag{13}$$

Experimental results conducted in the next section show that these four RR metrics can measure the image quality of a spatial resolution enhanced image very well. Even though these four RR metrics are derived for a special spatial resolution enhancement factor 2×2, it is easy to extend it to other spatial resolution enhancement factor $M\times N$, where both M and N are positive integers.

The IBP is chosen to enhance the spatial resolution of the testing images. For simplicity, this paper only considers spatially increasing the resolution by a factor of 2×2. It is easy to extend IBP to even higher resolution enhancement. IBP consists of two

steps: (i) projection, and (ii) back-projection. It enhances spatial resolution of an image by performing projection and back-projection iteratively until satisfactory results are obtained. In IBP, the imaging is regarded as a projecting process that includes shifting, under-sampling and blurring operations to generate a set of low resolution images. So the reconstruction of a high-resolution image from these low resolution images can then be regarded as a back-projecting process which includes de-blurring, up-sampling and anti-shifting operations. This back-projection is performed in an iterative way. The IBP algorithm converges rapidly, and can meet the need of real-time processing since it only deals with some simple operations. Generally, the resultant image has satisfactory visual effect after 10 iterations.

For the sake of comparison with IBP, interpolation is used to enhance the spatial resolution of the testing images. The bilinear interpolation was chosen as the interpolation method in the experiments.

3. EXPERIMENTAL RESULTS

In this section, a number of experiments were conducted to demonstrate the feasibility of the proposed RR metrics. Three hyperspectral data cubes were tested in this paper. The 2-dimensional (2D) band images of the data cubes are used to test the proposed RR metrics. The first hyperspectral data cube was acquired using the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) in the Cuprite mining district, Nevada, in 1997. The original scene with size of 614×512 pixels and 224 bands is available online at http://aviris.jpl.nasa.gov/html/aviris.freedata.html. The upper-right corner of the scene that consists of 350×350 pixels and 224 bands are cut out. This scene is well understood mineralogically and it has been made a standard test site for validation and assessment of remote sensing methods (Chen and Qian, 2007, 2008a, 2008b, 2009; Wang and Chang, 2006). Due to water absorption and low SNR, the bands 1-3, 105-115 and 150-170 are removed. As a result, a total of 189 bands are used in our experiments. Figure 1 shows the image of the Cuprite data cube at wavelength 827 nm (band #50).

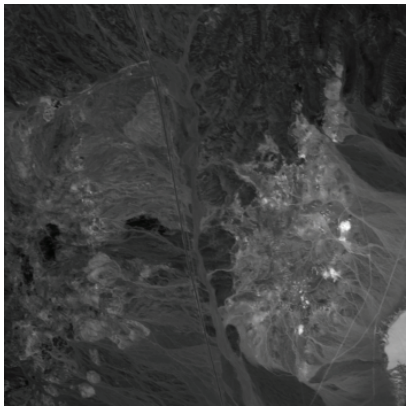


Figure 1. The AVIRIS Cuprite scene at displayed at wavelength 827nm (spectral band #50).

The second hyperspectral data cube was acquired using the airborne Short-wave-infrared Full Spectrum Imager II (SFSI-II). The data cube was collected over Key Lake in northern Saskatchewan, Canada for studying the capability of imaging spectrometers in identifying uranium mine and associated activities. The data cube was acquired with a ground sample distance (GSD) of 3.19m×3.13m. The size of the data cube is 1090 lines by 496 pixels by 240 bands. The scene of the testing data cube includes a mill complex and a mine complex. Figure 2 shows an image at wavelength 1304 nm (band #16) of this data cube.

The third hyperspectral data cube was also collected using the SFSI-II for studying target detection from short wave infrared hyperspectral imagery. The GSD of the data cube is 2.20m×1.85m. The size of the data cube is 140 lines by 496 pixels by 240 bands. Man-made targets with different materials and sizes were deployed in a mixed of sand and low-density grass cover within the scene of the data cube. Seven pieces of awnings with varying sizes ranging from 12m×12m to 0.2m×0.2m, four pieces of polythene, four pieces of white tarp and four pieces of white cotton with varying size ranging from 6m×6m to 0.5m×0.5m were deployed. In addition, a 3m×3m piece of white tarp was placed on a large vinyl turf mat of size 11m×14m. Figure 3 shows a region-of-interest (size: 140×140) at wavelength 1289 nm (band #13) of this data cube.

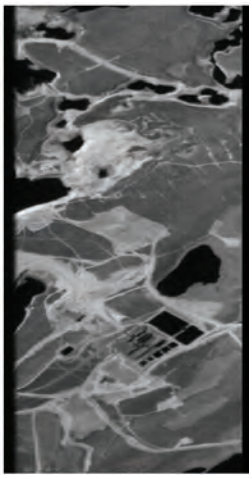


Figure 2. The Key Lake scene displayed at wavelength 1304 nm (spectral band #16).

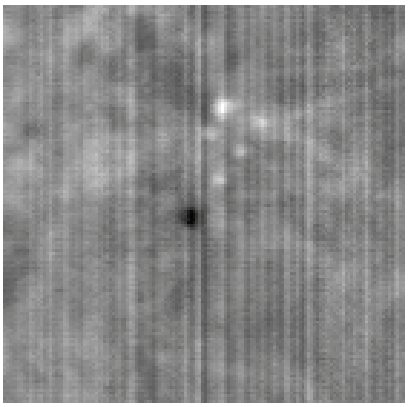


Figure 3. The SFSI-II data cube with man-made targets displayed at wavelength 1289 nm (spectral band #13).

Ideally, the IBP and bilinear interpolation should be used to enhance the spatial resolution of every spectral band in the

hyperspectral data cubes, and then the proposed four RR metrics were used to assess their image quality. However, due to the workload of computing the IBP and the metrics for the whole datacubes, as a consequence, bands #16, #50 and #13 have been chosen in the experiments for the Cuprite data cube, the Key Lake data cube, and the Target data cube, respectively.

The PSNR, Q index, MSSIM, and VIF are of widely used FR metrics in image processing. In the experiments, these four FR metrics are compared to their corresponding RR metrics. Table 4 lists the experimental results of the metrics applied to the spatial resolution enhanced images by using IBP and interpolation. The IBP was run for 30 iterations in order to generate a higher quality spatial resolution enhanced image.

Data cube	Spatial Enhancement Method	Full-Reference metrics				Reduced-Reference metrics			
		PSNR	Q	MSSIM	VIF	PSNR	Q	MSSIM	VIF
Cuprite	IBP	36.51	0.82	0.91	0.69	43.82	0.97	0.99	0.87
	Interpolation	35.67	0.78	0.90	0.48	38.37	0.92	0.96	0.75
Key Lake	IBP	34.87	0.75	0.89	0.77	40.11	0.96	0.98	0.87
	Interpolation	32.41	0.70	0.87	0.54	34.59	0.87	0.94	0.78
Target datacube	IBP	53.33	0.78	0.99	0.77	61.67	0.97	1.00	0.98
	Interpolation	53.14	0.74	0.99	0.65	56.35	0.89	1.00	0.88

Table 4. Experimental results of four FR image quality metrics and the four proposed RR metrics of the test images that are spatially enhanced by using the IBP and interpolation methods. For the FR metrics, a test image is first down-sampled at a factor of 2×2, then spatially enhanced at a factor of 2×2. For the proposed RR metrics, a test image is spatially enhanced by a factor of 2×2 without a prior down-sampling.

For the FR metrics, a test image is first down-sampled at a factor of 2×2, then is spatially enhanced at a factor of 2×2 in order to satisfy the requirement of the processed image having the same size as the reference image. For the proposed RR metrics, an original test image is spatially enhanced at a factor of 2×2 without a prior down-sampling. From the table, it can be seen that IBP-based method always produces better results than the bilinear interpolation no matter whether the original image is down-sampled or not. More importantly, the proposed RR metrics measure the image quality of the spatial resolution enhanced images very well, and they are consistent with the corresponding FR metrics. This indicates that the proposed RR metrics are reliable metrics for measuring the quality of the spatial resolution enhanced images.

4. CONCLUSION

In this paper, four new RR metrics are proposed to measure the quality of the spatially enhanced hyperspectral images. These metrics do not require the sizes of the reference and test images to be the same. However, all FR metrics published in the literature require both images to have the same size. The IBP and bilinear interpolation are used to increase the spatial resolution of a testing image. Experimental results show that the proposed four RR metrics can measure the image quality of the spatial resolution enhanced images very well. Even though only hyperspectral images are tested in this paper, the proposed metrics can be used to measure the image quality of any other spatially enhanced images.

REFERENCES

Chen, G. Y. and Qian, S. E., 2007. Dimensionality reduction of hyperspectral imagery using improved locally linear embedding. *Journal of Applied Remote Sensing*, 1, pp. 013509.

Chen, G. Y. and Qian, S. E., 2008a. Evaluation and comparison of dimensionality reduction methods and band selection. *Canadian Journal of Remote Sensing*, 34(1), pp. 26-32.

Chen, G. Y. and Qian, S. E., 2008b. Simultaneous dimensionality reduction and denoising of hyperspectral imagery using bivariate wavelet shrinking and principal component analysis. *Canadian Journal of Remote Sensing*, 34(5), pp. 447-454.

Chen, G. Y. and Qian, S. E., 2009. Denoising and dimensionality reduction of hyperspectral imagery using wavelet packets, neighbour shrinking and principal component analysis. *International Journal of Remote Sensing*, 30(18), pp. 4889-4895.

Engelkea, U., Kusuma, M., Zepernick, H. J. and Caldera, M., 2009. Reduced-reference metric design for objective perceptual quality assessment in wireless imaging. *Signal Processing: Image Communication*, 24(7), pp. 525-547.

Irani, M. and Peleg, S., 1991. Improving resolution by image registration. *Computer Vision, Graphics and Image Processing*, 53, pp. 231-239.

Irani, M. and Peleg, S., 1993. Motion analysis for image enhancement: resolution, occlusion and transparency. *Journal of Visual Communications and Image Representation*, 4, pp. 324-335.

Li, Q. and Wang, Z., 2009. Reduced-reference image quality assessment using divisive normalization-based image representation. *IEEE Journal of Selected Topics in Signal Processing, Special issue on Visual Media Quality Assessment*, 3(2), pp. 202-211.

Sheikh, H. R., Bovik, A. C. and Cormack, L., 2005. No-reference quality assessment using natural scene statistics: JPEG2000. *IEEE Transactions on Image Processing*, 14(11), pp. 1918-1927.

Sheikh, H. R. and Bovik, A. C., 2006. Image information and visual quality. *IEEE Transactions on Image Processing*, 15(2), pp. 430-444.

Wang, Z. and Bovik, A. C., 2002. A universal image quality index. *IEEE Signal Processing Letters*, 9, pp. 81-84.

Wang, Z., Bovik, A. C., Sheikh, H. R. and Simoncelli, E. P., 2004. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), pp. 600-612.

Wang, J. and Chang, C.-I., 2006. Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 44(6), pp. 1586-1600.

Wang, Z. and Simoncelli, E. P., 2005. Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. *Human Vision and Electronic Imaging X, Proc. SPIE*, vol. 5666, San Jose, CA.

ACKNOWLEDGEMENTS

This work was partially supported by the Visiting Fellowship in Canadian Government Laboratory coordinated by the Natural Sciences and Engineering Research Council of Canada (NSERC). The authors would like to thank J. Levesque at Defence Research and Development Canada for providing the test datacubes.

A COMPARATIVE CASE OF STUDY OF IMAGE SHARPENING

V. F. Rodríguez-Galiano [a, *], E. Pardo-Igúzquiza [a], M. Chica-Olmo [a], M.J. García-Soldado [a], J. P. Rigol-Sánchez [b], M. Chica-Rivas[c]

[a] Dept. of Geodynamics, Science Faculty, 18071 Granada, Spain - (vrgaliano, mchica)@ugr.es
[b] Dept. of Geology, University of Jaén, Campus Las Lagunillas s.n., 23071 Jaén, Spain - jprigol@ujaen.es
[c] Dept. of Mathematic Analysis, Science Faculty, 18071 Granada, Spain

KEY WORDS: Statistics, Fusion, Algorithms, Landsat, Spatial

ABSTRACT:

In this paper we present a comparative case of study of different methodologies for image sharpening. The evaluated methodologies are classic procedures such as Brovey (BR), Intensity Hue Saturation (FIHS), and Principal Component Analysis (PCA); two procedures based on wavelet transforms: Wavelet à Trous (WAT) and MultiDirection MultiResolution (MDMR); and one method of a geostatistical nature, Downscaling Cokriging (DCK). The comparison of the fused images is based on the quantitative evaluation of their spatial and spectral characteristics by calculating statistical indexes and parameters to measure the quality and coherence of the resulting images.
Synthesis of the obtained results shows that the algorithm fusions based on wavelet and DCK yielded better results than did the classical algorithms. Particularly, the DCK geostatistical method does not introduce artefacts in the estimation of the digital levels corresponding with the source multispectral image and, in this sense, can be considered the most coherent method. The MDMR method produces the merged images with the highest spatial quality.

1. INTRODUCTION

The arrival of new sensors and satellites in recent decades has notoriously favoured the availability of remotely sensed images with enhanced spatial and spectral resolution. In practice, a more effective use of this information entails the application of image fusion techniques in order to obtain a final product with improved spatial and spectral resolution suitable for a specific application.

The simplest fusion techniques are based on the direct substitution of some bands for visualization or in a simple arithmetic transformation, such as the Brovey (BR) method or the FastIHS method (Tu, 2005). Some other classic image fusion algorithms are more complex and involve transformations of the images and substitution of components, e.g. Principal Component Analysis (PCA) or Intensity Hue Sturation (IHS) transform.

A set of image fusion methods based on wavelet theory have recently been proposed (Amolins et al., 2007). These fusion algorithms may be considered as an extension of the High Pass Filtering (HPF) method, since they hold that spatial information is contained in the high frequencies. The wavelet transforms extract detailed information of the panchromatic image to integrate it in the multispectral image by means of methods based on the frequency or the spatial context. An advantage of these techniques is that the wavelet function can be modified to enhance specific features, which can be useful for a particular application (Amolins et al., 2007).

A further methodological alternative to the above is founded on geostatistical methods, which explicitly account for spatial variability characteristics of the images to be fused (Chica-Olmo and Abarca-Hernández, 1998). The geostatistical fusion model is based on the Cokriging method. One variant of this methodology is the Downscaling Cokriging method (DCK), proposed by Pardo-Iguzquiza et al. (2006), which considers relevant aspects for image fusion purpose such as pixel size (information support), the direct and cross-spatial correlations of the image digital values, and the point spread functions of the sensors.

Although several comparative studies of remote sensing image fusion methods have been published, there are only a few that include a detailed assessment of results obtained with a broad range of available techniques. The aim of this study is to perform a detailed comparative analysis of a set of image-fusion algorithms representative of the different methodological approaches. To this end, several classic methods based on arithmetic transformations or substitution of components were chosen: Brovey (BR), Fast IHS (FIHS) and Principal Component Analysis (PCA); two methods based on wavelet transforms: Wavelet à Trous (WAT) and MultiResolution MultiDirection (MDMR); and finally, a method of a geostatistic nature, Downscaling Cokriging (DCK).

2. RELATED WORK

2.1 Image sharpening approaches

Below we briefly describe the selected fusion algorithms that were chosen for comparative study.

BR Method: It is a very popular method of easy application, which is based on simple arithmetic applications, in which each band of the colour image is multiplied by the high resolution image and divided by the sum of the multispectral bands.

PCA Method: PCA is based on the application of a classic procedure of principal component analysis of the original bands

* Corresponding author

of the mustispectral image. In the calculation of the principal components, the common information of the set of multispectral bands is contained, mainly, in the first component. This component is substituted by the panchromatic band, equivalent in radiometric information content, but having better spatial resolution. Inverse transformation allows the fused image to be obtained.

FIHS Method: The IHS is based on the transformation of the colour space, from RGB to IHS, and substitution of the resulting band intensity with the panchromatic image of high spatial resolution. By applying the inverse transformation after substitution, one obtains a multispectral image that is similar to the initial one, but has improved spatial resolution. The FIHS fusion algorithm is based on the same theoretical principals as the IHS, but the process of inverse transformation is simplified (Tu, 2005).

WAT Method: Wavelet transforms are considered as a bank of filters that, upon application to a sequence of levels of decomposition, divide the signal (e.g. satellite image) into high and low frequency components (Amolins et al., 2007). When decomposition at different levels is applied, we speak of multiresolution decomposition.

The transform denominated Wavelet à Trous, or WAT consists basically of the application of a series of consecutive convolutions for different levels of degradation. WAT calls for an iterative filtering process, in which a series of degradation filters are used to obtain the wavelet. Because it is not a decimal algrorithm (with holes), the point of departure is an initial filter to which rows and columns are iteratively added, with zeros introduced between the rows and columns of the filter of the previous iteration, until the desired resolution is achieved.

The WAT method, unlike algorithms such as the pyramidal one of Mallat is characterized by the directional independence of the filtering process, without spatial compression of the different levels of degradation. Therefore, the image for each level of degradation has half the resolution of the previous one, but the same size, so that the information contained in each is redundant.

The wavelet coefficients $A^{kj}_{uj+n}(x)$ are calculated as the difference between two consecutive levels of degradation:

$$A^{k_j}_{u_j+n}(x) = DN^{k_j}_{u_j+n-1}(x) - DN^{k_j}_{u_j+n}(x) \quad (1)$$

DN represents the digital number of a pixel of location x = (x,y) belonging to spectral band k_j of the original image.

Following an additive criterion, if $DN^{kj}_{uj+n}(x)$ represents the successive degradations that contain the information of low frequencies of the original multispectral image, and $A^{kj}uj+n(x)$ the respective wavelet coefficients that contain the high frequency information, it is possible to obtain a fused image of high resolution by means of the sum of the low frequencies contained in the degraded multispectral image and the high frequencies extracted from the panchromatic image.

$$\widehat{DN}^{k_0}_{u_0}(x) = DN^{k_j}_{u_j+n}(x) + \sum_{k=1}^{n} A^{k_j}_{u_{j+n}}(x) \quad (2)$$

MDMR Method: The MultiDirection MultiResolution fusion algorithm (Lillo-Saavedra and Gonzalo, 2007) is a modification of the WAT that incorporates directional transforms. It is an algorithm meant to attain optimal equilibrium between the spectral and the spatial resolution of combined images, via the application of directional ellipsoidal filters.

The fusion process is virtually identical to that explained under the WAT method.

$$A^{k_j}_{\theta_n}(x) = DN^{k_j}_{\theta_n}(x) - DN^{k_j}_{\theta_n-1}(x)$$

$$\widehat{DN}^{k_0}_{u_0}(x) = DN^{k_j}_{\theta_l}(x) + \sum_{k=1}^{l} A^{k_j}_{\theta_n}(x) \quad (3)$$

However, we see that the level of degradation has been replaced by that of the directional filter of orientation θ. Unlike WAT, this is a highly anisotropic algorithm, which allows for trade-off between the desired spatial and spectral resolutions (see Lakshmanan 2004).

Downscaling Cokriging Method: The fused image of high spatial resolution obtained by means of this geostatistial method, DCK, is expressed as a linear combination of the experimental images (Pardo-Iguzquiza et al., 2006; Atkinson et al., 2008):

$$\widehat{DN}^{k_0}_{u_0}(\mathbf{x}_0) = \sum_{j=1}^{M}\sum_{i=1}^{n_j} \lambda^0_{ji} DN^{k_j}_{u_j}(\mathbf{x}_i) \quad (4)$$

where:
DN^b_a represents the digital number of a satellite image for the spectral band b and with a spatial resolution (pixel size) a and for a particular spatial location. The circumflex symbol above DN denotes that it is an estimated image or one fused by cokriging, whereas without the accent it is designated as an experimental image. Other annotations are:
$b=k_0$ spectral band whose spatial resolution should be improved.
$b=k_j$ experimental spectral band included in the process of fusion by cokriging.
$a=u_0$ spatial resolution or pixel size of the fused image.
$a=u_j$ spatial resolution or pixel size of an experimental image used in the fusion.
M: number of experimental bands used in the fusion.
n_j: number of pixels of the neighborhood used for the experimental image of the spectral band.
λ^0_{ji} : optimal weight applied to $DN^{kj}_{uj}(x_i)$ in the estimation of $DN^{k0}_{u0}(x_i)$.
The optimal weights given above are obtained by means of the resolution of a system of linear equations known as a cokriging system. This system is derived by imposing that the estimator be unbiased:

$$\mathrm{E}\{\widehat{DN}_{u_0}^{k_0}(\mathbf{x}_0) - DN_{u_0}^{k_0}(\mathbf{x}_0)\} = 0 \quad (5)$$

and minimizing the variance of estimation

$$\mathrm{E}\{[\widehat{DN}_{u_0}^{k_0}(\mathbf{x}_0) - DN_{u_0}^{k_0}(\mathbf{x}_0)]^2\} \rightarrow \min \quad (6)$$

where E{.} is the operator of mathematical expectation.

The cokriging system also accounts for three key aspects of fusion: the size of the pixel of the experimental images (support effect), the direct and crossed variograms of the radiometric bands, and the point spread functions of the sensor. (For a more detailed description of the cokriging system see Pardo-Iguzquiza et al., 2006; Atkinson et al., 2008).

2.2 Evaluation approaches

A set of statistical parameters and indexes were calculated to quantify the differences between the spectral information of the compared images, and, moreover, to measure the spatial and spectral quality overall:

-Correlation coefficient between the original multispectral image and the fused images.

-Mean Error and Root Mean Square Error of the original and the fused image.

-The ERGAS index (Erreur Relative Globale Adimensionnelle de Synthèse) (Wald 2000):

$$ERGAS_{spectral} = 100\frac{h}{l}\sqrt{\frac{1}{N}\sum_{k=1}^{N}\frac{\left(RMSE_{spectral}(B_k)\right)^2}{M_k^2}} \quad (7)$$

where h/l is the ratio between the resolution of the panchromatic image and the multispectral image, N is the number of spectral bands (B_k) of the fused image, M_k is the mean value of each spectral band, and RMSE is the Root Mean Square Error calculated between the fused image and the multispectral original.

To measure the spatial quality of fused images, authors Lillo-Saavedra et al. (2005) put forth a modification of the classic spectral ERGAS, referred to as the spatial ERGAS:

$$ERGAS_{spatial} = 100\frac{h}{l}\sqrt{\frac{1}{N}\sum_{k=1}^{N}\frac{\left(RMSE_{spatial}(B_k)\right)^2}{P_k^2}} \quad (8)$$

This index differs from the previous one in that it uses the original panchromatic band (P_K) instead of the multispectral one.

-The Image Quality Index, proposed by Wang and Bovik (2002) as an alternative to the Mean Square Error. It models the differences between two given monochromatic images as a combination of three separate factors: loss of correlation, luminance distortion, and contrast distortion.

$$Q = \frac{4\sigma_{OF}\bar{O}\bar{F}}{\left(\sigma_O^2 + \sigma_F^2\right)\left[\left(\bar{O}\right)^2 + \left(\bar{F}\right)^2\right]} \quad (9)$$

where and are the mean of the original and the fused images, respectively, σ^2_O and σ^2_F are the variances, and σ_{OF} is the covariance between the original image and the fused one.

- Index of spatial quality proposed by Zhou et al. (1998). This index measures the spatial quality of a fused image in relation with the spatial information provided by the panchromatic image. The algorithm applies a Laplacian filter to extract the high frequency information and compute the correlation coefficient between the sharpened image and the original panchromatic one.

-The "coherence measure" between the fused images and reality was obtained by calculating the correlation coefficient and the real errors (Mean error, ME and Root Mean Square Error, RMSE; (see Figure 2).

3. RESULTS AND DISCUSSION

The study is illustrated using a sector of a multispectral Landsat7 ETM+ scene of 944 km^2 (1024x1024) and its corresponding panchromatic image, with a spatial resolution of 30m and 15m, respectively. The image was acquired on 20 July 2002 over the metropolitan area of Granada, in southeast Spain. The scene corresponds to path 200 row 34 of the Landsat Worldwide Reference System (WRS) (figure 1).

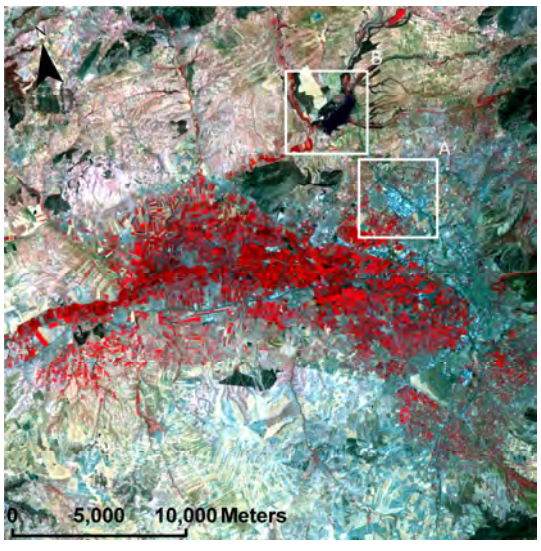


Figure 1. False colour composition image of the study area. Boxes A and B are two sectors of different land cover context.

The best reference for assessing the quality of a fused image is obviously the "true image" that the analyst wishes to obtain via the fusion method. In practice, this is however not feasible. For this reason, we designed an experiment in which the original

multispectral and panchromatic images could be degraded (by a factor of 2) to resolutions of 120m and 30m, respectively, in order to obtain fused images with a resolution of 30m (figure 2). In this way we were able to compare the results of fusion with the "true" or "real" image (e.g., coherence measures).

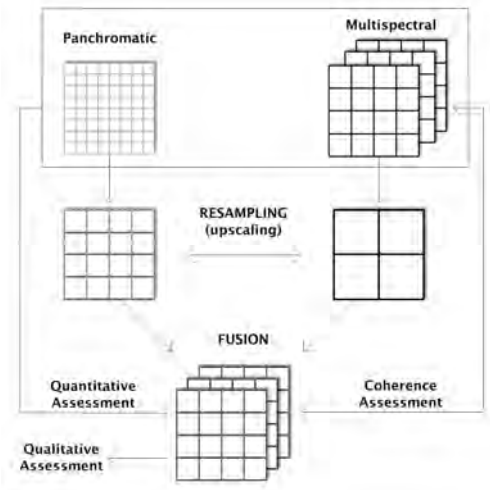


Figure 2. General scheme of the methodology used for comparative assessment of image fusion algorithms.

3.1 Application of the fusion algorithms

Classic methods: The classic fusion algorithms, BR, PCA or FIHS. These methods do not require definition of filters or the study of spatial variability between images, as required in the case of wavelet based and geostatistical based procedures.

Wavelet based:
WAT
In order to apply Wavelet à Trous fusion, a fusion ratio of 4:1 between the degraded multispectral Landsat image (120 m) and the degraded panchromatic image (30m) was considered. Two levels of degradation were applied to the multispectral image, so that two sets of wavelet coefficients were obtained, one containing detail between 120 m and 60 m and the other from 60 m to 30 m.

MDMR
In order to establish directionality and the optimal filter parameters, a great number of experiments were carried out applying different levels of degradation (l = 21, 22, 23, 24) for different combinations of a and b. The values of the filter parameters (a and b) were divided into two intervals. The first, defined between 0.1 and 0.5, using intervals of 0.1; and the second was ranged from 1 to 5 at intervals of 1. This gave a total of 100 different combinations for each partition frequency or degradation level (400 fused images). The resulting products of fusion were evaluated quantitatively using the ERGAS spatial and spectral indexes. For the selection of the best fused image by means of the MDMR algorithm, we determined the one in which the mean spatial and spectral ERGAS were lowest, and the difference between the two close to 0. This served as a guarantee of quality of the fusion, while affording balance between the spatial and spectral resolution of the fused image. According to the restrictions explained above, a filter with four directions and an adjustment of parameters a and b was selected which provided a fused image with a mean ERGAS equal to 2.15. Nonetheless, the values of a and b, together with the number of directional filters, can be adjusted to highlight the spectral or the spatial resolution so as to attain lower ERGAS values.

Geoestatistical based (DCK): This method requires the variographic analysis of the multispectral and panchromatic images: the experimental and the induced models of the simple variograms of the different bands of the multispectral and panchromatic images, as well as the cross-variograms between these images. A linear model of corregionalization with two superimposed exponential structures was used: one of short range (45 m) and the other of long range (728 m). The practical ranges are 135 m and 2184 m, respectively. The sills of the simple and cross-variograms of the multispectral and panchromatic bands at point support level were all calculated using a process of numeric deconvolution and an adjustment of weighted squared minima.

Fusion by downscaling cokriging was done using two bands, the band whose spatial resolution we wished to improve, and the panchromatic one. The results of the Cokriging system provided the weights that were applied to the high and low spatial resolution images; that is, the multispectral and panchromatic ones, respectively.

3.2 Evaluation of the overall quality

In this section, we present the results of the fusions and the assessment of the spectral and spatial quality of the fused images.

Two subsectors with different environmental context were chosen for detailed evaluation (figures 1 and 3). The fused images show considerable differences in visual quality depending on the integration technique applied. The BR and PCA methods have a negative impact on the colour of the image, decreasing contrast and increasing colour saturation. The FIHS and the WT methods achieve better spatial detail and give rise to sharper images, reflected most notably in the linear features present in the urban zones (figure 3 sector A). However, the greater the spatial enhancement the greater the spectral distortions and moreover, contrast is reduced, and an effect of radiometric homogenization is produced, which causes a loss of texture. This effect is particularly appreciable in the image obtained using the MDMR method, where the vegetation areas seem to be fuzzy. Finally, the geostatistical method (DCK) is the one that best conserves contrast, saturation and texture with respect to the original reference image, yet in certain areas, for instance in the urban ones, a coarser spatial resolution can be observed.

In order to quantify the quality of the images several indexes were calculated taking into account the original multispectral Landsat image, in the case of the spectral indexes, and the panchromatic image, in the case of the spatial indexes.

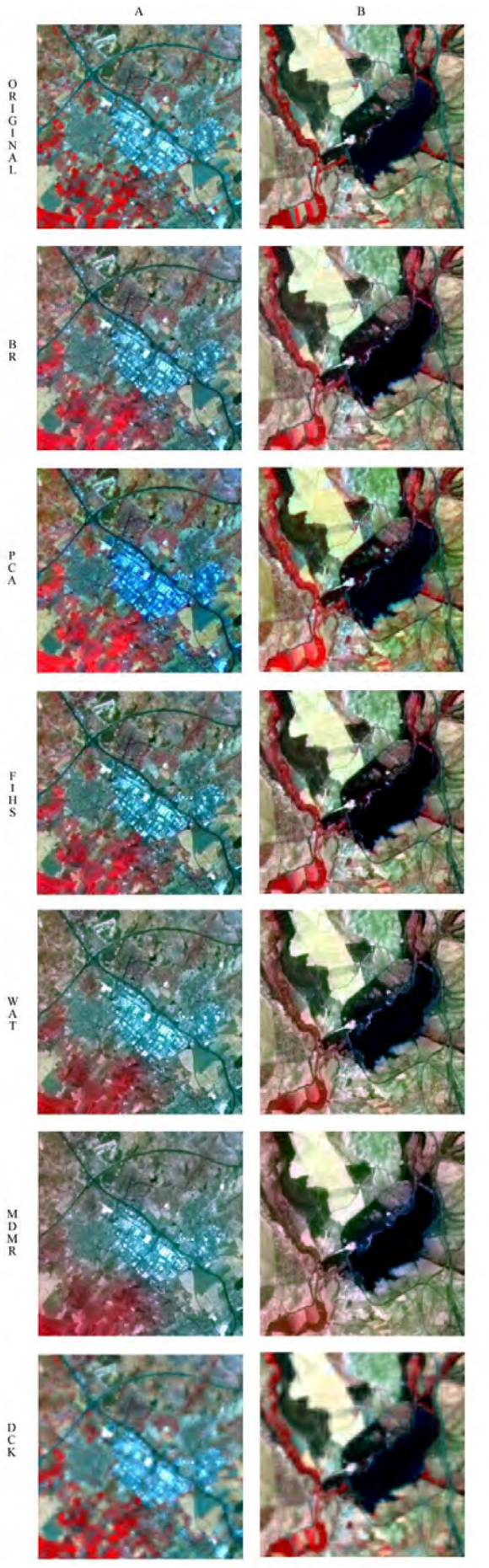


Figure 3. False colour compositions 432 (RGB) of the original multispectral image and the fused images for test sectors A and B.

In general, correlation coefficients indicate that a relatively large correlation in the near infrared (NIR) band exists for all the fusion methods analysed (0.896 to 0.958) (table 1). Correlations are however considerably larger for the green (0.940 to 0.962) and red visible (0.952 to 0.967) bands, due to a better spectral correspondence with the panchromatic band. With regard to image integration methods performance, PCA presents the smallest correlation coefficients for all the bands (0.896 to 0.955). In contrast, the DCK offers the best correlation with the multispectral image (0.958 to 0.967). Table 1 show that the RMSE are small for methods DCK, WAT and MDMR, while the rest of the algorithms present larger RMSE values, especially the BR method.

	BR	PCA	FIHS	WAT	MDMR	DCK
CC G	0.953	0.943	0.940	0.946	0.944	0.962
CC R	0.967	0.955	0.966	0.958	0.952	0.967
CC NIR	0.931	0.896	0.945	0.932	0.928	0.958
RMSE G	51.68	17.05	15.14	5.82	5.90	4.88
RMSE R	66.44	24.18	17.39	8.40	9.11	7.49
RMSE NIR	65.83	23.53	12.60	6.58	6.75	5.16
Spatial ERGAS	10.327	3.005	2.826	2.431	2.129	2.943
Spectral ERGAS	11.036	3.284	2.845	2.111	2.167	1.319
Average ERGAS	10.682	3.144	2.835	2.271	2.148	2.131
Q 8x8	0.647	0.875	0.881	0.871	0.860	0.879
Q 16x16	0.658	0.901	0.911	0.905	0.887	0.921
Q 32x16	0.665	0.914	0.928	0.925	0.907	0.940
Q 64x64	0.668	0.922	0.939	0.938	0.922	0.951
Q 128x128	0.670	0.926	0.947	0.948	0.935	0.960
Average Q	0.662	0.907	0.921	0.917	0.902	0.930
Zhou	0.972	0.972	0.995	0.997	0.996	0.857

Table 1. Values of the different parameters analysed to estimate the spectral and spatial quality of the fused images.

With regard to the spatial, spectral and mean ERGAS values (table 1), all the fusion methods, except the BR and PCA, generate good quality merged images. However, image fusion methods based on wavelet transforms (MDMR and WAT) and geostatistics (DCK) clearly outperform the rest of algoritms. MDMR and DCK are the ones providing larger spatial and spectral quality (2.129 and 1.319 respectively). DCK has a lower mean ERGAS than the others with a value equal to 2.131. The WAT method presents indexes of spatial and spectral quality that are better balanced (2.431 and 2.111 respectively). All the fusion methods, except BR, result in improved spectral quality with respect to the degraded multispectral image. The

DCK method yields the best results with an average Q equal to 0.930.

The Zhou spatial index (table 1) present high values in all cases, except under DCK, which gives a value of 0.857.

3.3 Assessment of coherence

The coherence of digital levels when comparing the target image and those estimated by means of the fusion algorithm has been considered in this study. We elaborated a "coherence measure" based on: mean error, Root Mean Square Error (table 2) and the correlation coefficient of each band estimated with respect to its corresponding "true" multispectral band (table 2). DCK is the most coherent, as it presents a maximum correlation coefficient (practically equal to 1) and it minimizes the RMSE for all the bands. The rest of the methods give correlation coefficients that are similar (all lower), whereas the RMSE and the ME of the classic methods are significantly less coherent than those of the wavelet methods.

	BR	PCA	FIHS	WAT	MDMR	DCK
ME G	-50.15	-15.73	-13.73	0.03	0.0341	-0.44
ME R	-63.25	-21.82	-15.69	0.03	0.0378	-0.45
ME NIR	-64.69	20.90	-11.08	0.04	0.0387	-0.44
RMSE G	51.40	16.31	14.17	3.00	3.6581	0.66
RMSE R	65.89	23.20	16.10	4.88	5.8669	0.69
RMSE NIR	65.59	22.98	11.45	3.65	4.3800	0.67
R G	0.98	0.97	0.97	0.98	0.9749	0.99
R R	0.99	0.98	0.99	0.98	0.9778	0.99
R NIR	0.97	0.92	0.98	0.97	0.9641	0.99

Table 2. Parameters of coherence between the true or real multispectral image and the fused images: mean error, Root Mean Square Error, and correlation coefficient.

4. CONCLUSIONS

The assessment of the global quality of all merged images has demonstrated that the algorithms based on wavelet transforms (WAT and MDMR) and the geostatistical algorithm, Downscaling Cokriging (DCK), produce better spectral and spatial results than the classic image fusion methods employed. These classic methods, with the exception of FIHS, introduce some colour distortions which can be observed in the visual analysis. The WT, along with FIHS method, enhance the spatial details of certain zones presenting specific patterns, such as the reticulate pattern of urban zones, although they introduce some distortions in more homogeneous zones such as areas covered with natural vegetation.
The assessment of the global quality of all merged images has demonstrated that the algorithms based on wavelet transforms and Downscaling Cokriging (DCK), produce better spectral and spatial results than the classic image fusion methods employed.
The analysis of the values of correlation coefficients, RMSE, spectral ERGAS and Q shows that the DCK method is the algorithm that best preserves the multispectral information of the original image. The MDMR method was the most efficient in increasing the spatial resolution of the image (as indicated by spatial ERGAS index and Zhou index). Finally, from the overall viewpoint of both spectral and spatial indexes, WAT is the method that presents the most balanced results.
The DCK is the most coherent method of those studied here, because it does not introduce artefacts in the estimation of the digital numbers.

REFERENCES

Amolins, K., Zhang, Y., and Dare, P. (2007). Wavelet based image fusion techniques -- An introduction, review and comparison. ISPRS Journal of Photogrammetry and Remote Sensing, 62, 249-263.

Atkinson, P.M., Pardo-Igúzquiza, E. and Chica-Olmo, M. (2008). Downscaling Cokriging for Super-Resolution Mapping of Continua in Remotely Sensed Images. IEEE Transactions on Geoscience and Remote Sensing, 46(2), 573-580.

Chica-Olmo, M., and Abarca-Hernandez, F. (1998). Radiometric coregionalization of Landsat TM and SPOT HRV images. International Journal of Remote Sensing, 19 (5), 997-1005.

Lakshmanan, V. (2004). A separable filter for directional smoothing. IEEE Geoscience and Remote Sensing Letters, 1(3), 192-195.

Lillo-Saavedra, M., Gonzalo, C., Arquero, A., and Martinez, E. (2005). Fusion of multispectral and panchromatic satellite sensor imagery based on tailored filtering in the Fourier domain. International Journal of Remote Sensing, 26, 1263-1268.

Pardo-Igúzquiza, E., and Chica-Olmo, M. (2006). Downscaling cokriging for image sharpening. Remote Sensing of Environment, 102(1-2), 86–98

Tu, T.M., Lee, Y.C., Chang, C.P., and Huang. P.S. (2005). Adjustable intensity-hue-saturation and BR transform fusion technique for IKONOS/QuiDCKBird imagery. Optical Engineering, 44(1).

Wald, L. (2000). Quality of high resolution synthesized images: is there a simple criterion?. International Conference on Fusion of Earth Data, France Nice, France: SEE GréCA.

Wang, Z., and Bovik, A. C. (2002). A universal image quality index, IEEE Signal Processing. Letters, 9(3), 81–84.

Zhou, J., Civco, D. L., and Silander, J. A. (1998). A wavelet method to merge Landsat TM and SPOT panchromatic data. International Journal of Remote Sensing, 19, 743-757.

ACKNOWLEDGEMENTS

We are grateful for the financial support given by the Spanish MICINN (Project CGL2006-06845/CLI) and Junta de Andalucía (Group RNM122).

MAPPING KEY HYDROLOGICAL VARIABLES USING REMOTELY SENSED IMAGES INTEGRATED IN A DISTRIBUTED WATER BALANCE MODEL

N. Sánchez [a, *], J. Martínez-Fernández [a], A. Calera [b], E. Torres [b], C. Pérez-Gutiérrez [a]

[a] CIALE, Centro Hispano Luso de Investigaciones Agrarias. Universidad de Salamanca. Río Duero, 12; 37185 Villamayor (Salamanca), Spain - (nilda, jmf, carpegu)@usal.es
[b] IDR, Instituto de Desarrollo Regional. Universidad de Castilla La Mancha. Campus Universitario s/n, 02071 Albacete. Spain - (Alfonso.Calera, Enrique.Torres)@uclm.es

KEY WORDS: Modelling, Imagery, Landsat, Calibration, Classification

ABSTRACT:

A cartographic application of remotely sensed images integrated in a distributed water balance is presented. The imagery consisted in an annual series of Landsat 7 ETM+, from where the NDVI (Normalized Difference Vegetation Index) and the land uses/land cover (lu/lc) map were extracted. The soil database came from a permanent experimental network of 23 stations for continuous measurement of soil moisture (REMEDHUS, Soil Moisture Measurement Stations Network) at the Duero Basin (Spain). The theoretical basis of the water balance application derives from FAO56, which is improved with a spatial frame implemented in the computerized tool HIDROMORE. In addition to the image inputs, climatic and soil properties databases were also integrated. Maps of Actual Evapotranspiration (AET), Deep Percolation (DP), and Irrigation rates (I) were extracted. Soil moisture series were simulated for each REMEDHUS station, and the validation of the simulation was performed comparing simulated soil moisture against field-measured. The results for the area-average soil moisture are: RMSE (Root Mean Square Error)=0.02 cm^3cm^{-3}, AI (Agreement Index)=0.90 and R (correlation coefficient)=0.82. It was found that the parameters which showed the most influence in the maps appearance were the soil characteristics (water content at field capacity and at wilting point), precisely the parameters that most spatially vary and are the most difficult to acquire.

1. INTRODUCTION

Combining remotely sensed data with a water balance is a frequent approach in hydrological modeling. Images can be used to estimate the spatial distribution of the evapotranspiration of locations within the scene as well as regional distributions of water balance components (Wegehenkel and Kersebaum, 2005). Regarding FAO56 (Allen et al., 1998), the use of remotely sensed NDVI can improve the tabulated values of some calculation parameters (i.e., Fraction of Vegetation Cover (FVC), basal crop coefficient (K_{cb})), and the lu/lc map can assign spatial patterns to some others, such as irrigation rates, root depth, and plant height. The innovation of HIDROMORE is that it makes possible the integration of standard calculations of FAO56 with the spatial database resulting from Landsat (for vegetation parameters), plus the soil map (for soil parameters) and the map interpolated with the available weather stations (for climatic parameters). The operational aim of this application −conceived for calculating irrigation optimum rates close to real time− rules out highly complex methods of image treatment. Regarding the geometric correction, an RMSE smaller than pixel size is suitable for a correct recording of all the images (Wolfe et al., 2002; Schroeder et al., 2006). As for the radiometric correction, due to the complexity of knowing the atmospheric parameters, the use of Dark Object Sustraction (DOS) method is frequent (Chavez, 1989). The main objective of this research was to present a cartographic application of multitemporal remotely sensed images integrated in a distributed water balance model to derive maps of key hydrological variables that are useful for water management. Simple but effective processing methods of calibration, correction and classification of the multitemporal remotely sensed images were applied in order to integrate the results in the balance calculation. A second goal was to discuss the resulting maps and validate them with in-situ soil moisture

2. STUDY AREA, SOIL AND CLIMATE DATABASES

The study area is located in the central region of the Duero Basin, in the Castilla y León region of Spain (Figure 1). The climate is continental-Mediterranean, with around 400 mm of average annual precipitation. Mean temperature is 12 ºC, with long, cold winters and hot summers. The average annual potential evapotranspiration is 1025 mm according to Penman-Monteith method. Land uses are mainly crops.

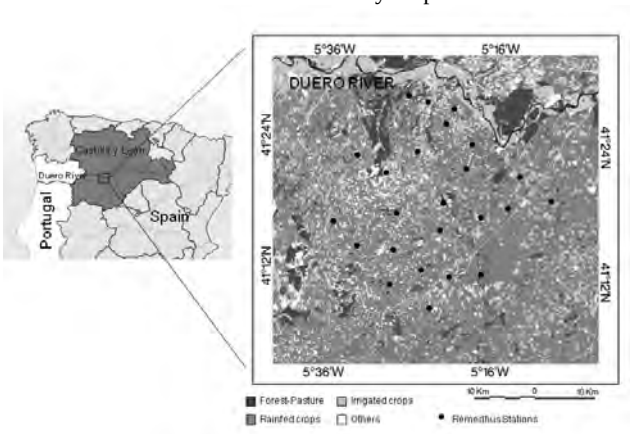


Figure 1. Lu/lc map of REMEDHUS area.

Distributed over 1300 km^2 the REMEDHUS network is made up of 23 stations for the measurement and monitoring of soil moisture (Martínez-Fernández and Ceballos, 2005). The soil database has a grid distribution of 3x3 km^2. For each one of these cells (146), several parameters are known at surface level: texture and water content at field capacity (θ_{FC}), at wilting point (θ_{WP}) and at saturation (θ_{sat}). The climatic database (precipitation, P, and Penmann-Monteith-based reference

* Corresponding author

evapo-transpiration, ET_0) came from one automatic weather station located in the REMEDHUS network.

3. MATERIAL AND METHODS

3.1 Image processing

A series of Landsat 7 ETM+ images from 2002 was used, scene 202/031, L1G path-oriented level. The series covered the whole year and allowed the study of the growing season of rainfed crops (spring) and irrigated crops (summer-autumn). All the scenes were clear and free of clouds.

For the geometric corrections, the physical model (Toutin, 2004) was chosen, based on the Landsat 7 orbital parameters and 24 ground control points, as well a digital elevation model of 1 m resolution, and the nearest neighbour resampling method. The global bundle adjustment showed an RMSE(X) of 0.68 pixels and RMSE(Y) of 0.70 pixels for all the images altogether.

The radiometric calibration and atmospheric correction were performed for red and infrared bands on each date. The procedure consisted of one-step transformation, following the method proposed by Pons and Solé-Sugrañés (1994). This method computes standard values of solar irradiance, calculated distance sun-earth, and standard values of aerosol optical depth for the red and infrared bands. DOS methods and topographic correction were also integrated.

Due to the difficulty of having a useful map during the study period, the spatial patterns of occupation of land are usually derived from remotely sensed maps. In this study, Landsat-derived map of lu/lc is used for: a) assigning the root depth, b) identifying irrigated crops, and c) thresholding the plant height and K_{cb} for each class. Irrigated crops map is used for estimating the irrigation water required to avoid water stress in their area along the growing season. Two factors were taken into account when selecting the thematic categories of the lu/lc map: similar temporal evolution of the plant cover (based on the evolution of NDVI curves) and representativeness of classes for the study area. For this last choice, Corine Land Cover map of 2000 and statistics of the Geographic Information System for Agricultural Plots (SIGPAC) were consulted (75% rainfed grasslands, 9% irrigated grasslands, 11% forest-pasture, 2% vineyards). The six categories chosen were rainfed cereal crops, irrigated crops, unproductive, water, vineyards, and forest-pasture (pasture cover with scattered trees).

The strategy of classification depends on modeling requirements. As NDVI is the most effective single spectral dimension to derive land cover types (Cihlar et al., 1996), its temporal series were made the basis of the lu/lc classification, and each NDVI date was an input band. Regarding the algorithm, two methods were tested: segmentation of NDVI values (Vincent and Pierre, 2003) and the maximum likelihood algorithm applied to a supervised classification. The training areas were extracted mixing the SIGPAC database and a field campaign.

For validating the two maps, control plot error and confusion matrix were applied. The map resulting of the supervised procedure had a mean accuracy of 86,83% and the map resulting of the segmentation, one of 69,2%. The confusion matrix showed a mean accuracy of 79,81% and kappa (Lillesand and Kiefer, 1999) of 0,69 for the supervised map and 78,85% and kappa 0,63 for the segmented map. In summary, the best procedure seemed the supervised classification, so the lu/lc map resulting of it was chosen for the integration in HIDROMORE.

3.2 Modelling

3.2.1 Theoretical background: FAO56 procedure, also called K_cET_0 approach, calculates reference and crop evapotranspiration from meteorological data and crop coefficients (Allen, 1998). Using the dual crop coefficient approach, the calculation of AET is performed as:

$$AET=ET_0(K_sK_{cb}+K_e) \quad (1)$$

In this dual form, the term $K_sK_{cb}ET_0$, represents the transpiration component, and the term K_eET_0, represents the soil evaporation component. Basal crop coefficient, K_{cb}, is the transpiration coefficient at a potential rate, i.e. when water is not limiting transpiration, and it is usually obtained from tabulated values; K_s describes the effect of water stress, and it is calculated according the water content in the root layer, that it is estimated from a daily water balance. The soil evaporation coefficient, K_e, is calculated from the water balance on the upper soil surface layer.

The soil moisture content in the root layer was calculated as a residual value of the water balance equation, and can be expressed as a water deficit or depletion (D_r). The balance was daily calculated in the following way, for day i:

$$D_{r,i}=D_{r,i-1}-P_i-I_i+AET_i+DP_i \quad (2)$$

where $D_{r,i}$ is the depletion for day i; $D_{r,i-1}$ is the depletion for the day i-1; DP_i is deep percolation for the day i, or the amount of water that exceeds the field capacity; and P_i, I_i and AET_i are effective precipitation, irrigation rate and AET for the day i respectively. Surface runoff was not considered in the current application.

The above described approach, which was developed initially for crops under optimal management conditions, can be also applied to natural or non-pristine vegetation (Allen et al., 1998). The problem in this case is how to obtain K_{cb} for crops or vegetation out of perfect growing conditions. The solution applied in this paper for solving this, is to obtain the time series of K_{cb} from the time series of NDVI, applying a previously developed relationship NDVI-K_{cb}. To obtain NDVI daily values, we interpolate linearly from the dates where the image is available. The schema of the application and other intermediate parameters are shown in Figure 2.

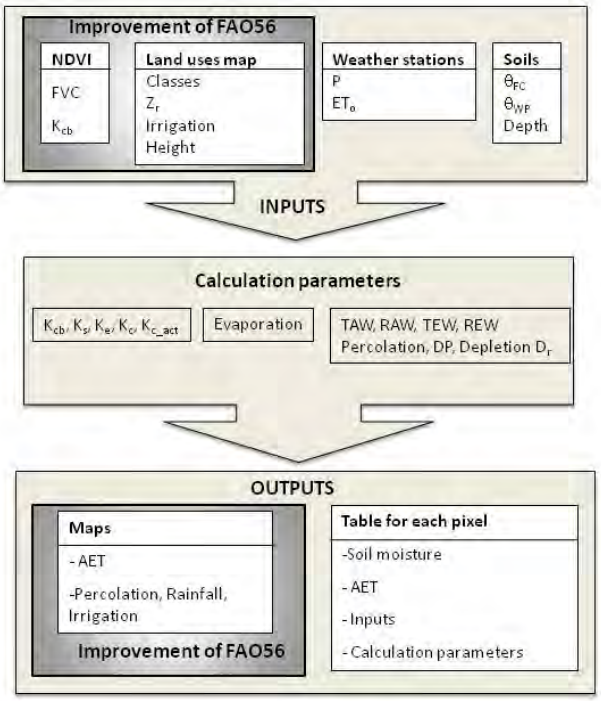


Figure 2. Schema of the FAO56 procedure in the HIDROMORE application. TAW, RAW, TEW, REW expressed water limits of the soil in the root zone (TAW and RAW) and at surface (TEW, REW). All other parameters are defined in the text.

3.2.2 HIDROMORE: HIDROMORE is a computerized tool for estimating evapo-transpiration and aquifer recharge based on FAO56 methodology in a distributed way. This application proposes an improvement of this methodology by incorporating the spatial databases (NDVI, lu/lc map, soil and climatic maps) and using them to parametrize the calculation. Thus, HIDROMORE transforms the water budget of FAO56 into a spatially distributed balance, and the resulting parameters can be studied as raster maps. The temporal scale (daily) and the distributed nature of the simulation (maps) make it adequate for a good management of water resources.

HIDROMORE uses the lu/lc map to compute some parameters of the model, as stated before. The NDVI has been used to calculate the K_{cb} according to the formula of Bausch and Neale (1987). NDVI was also the basis of the FVC calculation, following the observations on barley by González-Piqueras (2006).

$$K_{cb}=1.36\ NDVI-0.03 \quad (3)$$

$$FVC=1.19\ NDVI-0.16 \quad (4)$$

Both methods applied an empirical linear relationship. Even though the NDVI-K_{cb} equation can be different depending on the vegetation cover type, only the relationship for grassland was used for the entire area, owing to the prevalence of this cover (84%). Relationships for other crops such as vineyards are being studied for future research with in situ measurements.

4. RESULTS AND DISCUSSION

The distributed balance afforded a daily maps series of the results: AET, DP and I. Accumulated values can also be extracted (Figure 3). The map spatial sampling matches the Landsat spatial resolution, and for each pixel, the parameter takes a digital floating number corresponding to the parameter value. The influence of the soil database (tiles of 3x3km) is noticeable in the appearance of such maps.

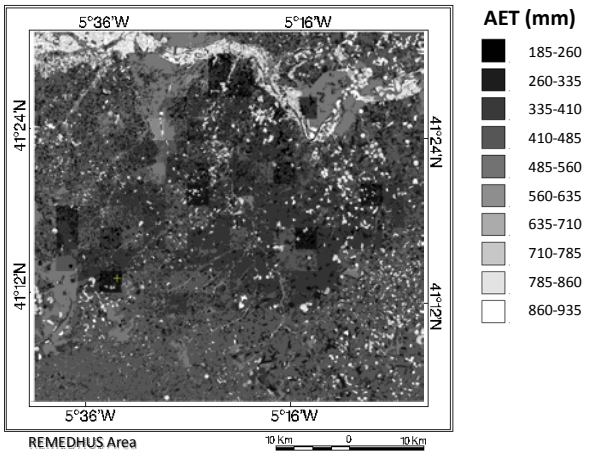


Figure 3. Map of accumulated AET in 2002.

Qualitative analysis was done for each class of the lu/lc map, extracting monthly and annual AET and DP (Table 4). The categories with higher evapotranspiration were irrigated crops and forest-pasture. The highest percolation was for irrigated crops and vineyards, due to their shorter growing period, together with more sandy soil texture in these categories. Irrigation maps can be used to extract the amount of theoretical irrigation water of each plot of this coverage. The mean irrigation amount in 2002 for the whole coverage was 458 mm (maximum is 656 mm and minimum 75 mm).

	AET (mm)	DP (mm)
Forest-pasture	533.30	9.00
Irrigated crops	809.13	183.44
Rainfed cereals	413.42	69.25
Vineyard	327.55	164.95

Table 4. Total AET and DP for lu/lc categories in 2002.

The monthly values showed that the highest evapotranspiration was for irrigated crops in summer –its growing period (Figure 5). For rainfed cereals, the full activity period was spring, when they consumed the natural precipitations and the water reserves of soil. For vineyard and forest-pasture, this period is delayed until the beginning of summer.

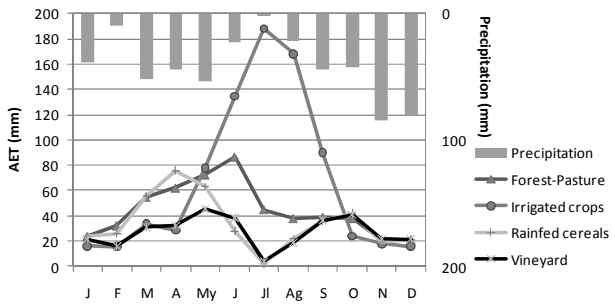


Figure 5. Montly AET for every lu/lc.

The quantitative map assessment was done by means of the simulated soil moisture against the field soil moisture at the REMEDHUS stations. Analysis of RMSE, R, and AI (Willmott, 1987) for 23 stations was performed. HIDROMORE afforded good results for the stations-average soil moisture (R=0.82; AI=0.90; RMSE=0.020 cm^3cm^{-3}), especially for rainfed cereals (R=0.91; AI=0.92; RMSE=0.024 cm^3cm^{-3}).

Regarding the maps, the gridded shape of them (Figure 3) shows the dependency of the water balance on the parameters of soils, particularly θ_{FC} and θ_{WP}. When these values were very low or very high for one cell, they afforded extremes values of evapotranspiration or percolation in this cell, producing a discontinuity.

5. CONCLUSIONS

Some results of a distributed water balance model based both on FAO56 and remotely sensed data were presented in this paper. The integration of the remotely sensed data afforded a detailed description of some vegetation parameters used in the calculation, otherwise these parameters should be tabulated under the FAO56 premises. The model outputs were maps of evapotranspiration, percolation and irrigation rates together with a soil moisture simulation over an area of 1300 km^2. Pre-processing of image consisted, despite more complex alternatives, on orthorectification by physical model, radiometric calibration and correction with standard conditions of atmosphere and DOS method, and topographic correction. Supervised classification had the best accuracy for extracting the lu/lc map. The results showed a good agreement between simulated and field soil moisture in terms of RMSE, AI and R. The maps qualitative analysis agrees with the expected behaviour of the agricultural uses in the area. The parameters that had more influence in the spatial distribution of the balance were θ_{FC} and θ_{WP}, related with the high spatial variability of soils. HIDROMORE improves FAO56 in integrating remotely sensed data and it affords maps of hydrological variables that are very useful for agricultural management. Future research will be done on simulated soil moisture maps *vs.* other image products of this parameter, i.e. SMOS Level 2 products.

REFERENCES

Allen, R.G., Pereira, L.S., Raes, D. and Smith, M., 1998. *Crop Evapotranspiration*. FAO Irrigation and Drainage, 56. Food and Agriculture Organization, Rome.

Bausch, W.C. and Neale, C.U., 1987. Crop coefficients derived from reflected canopy radiation: a concept. Transactions American Soc. Agric, 30(3), pp. 703-709.

Cihlar, J., Ly, H. and Xiao, Q., 1996. Land Cover Classification with AVHRR Multichannel Composites in Northern Environments *Remote Sens. Environ.*, 58, pp. 36-51.

Chavez, P.S., Jr., 1989. Radiometric calibration of Landsat Thematic Mapper multispectral images. *Photogramm. Eng. Rem. S.*, 55(9), pp. 1285-1294.

González-Piqueras, J., 2006. Evapotrans-piración de la Cubierta Vegetal Mediante la Determinación del Coeficiente de Cultivo por Teledetección. Extensión a Escala Regional: Acuífero 08-29 Mancha Oriental., Universitat de Valencia, Valencia, 334 pp.

Hu, C., Muller-Karger, F.E., Andrefouet, S. and Carder, K.L., 2001. Atmospheric correction and cross-calibration of LANDSAT-7/ETM+ imagery over aquatic environments: A multiplatform approach using SeaWiFS/MODIS. *Remote Sens. Environ.*, 78(1-3), pp. 99-107.

Lillesand, T.M. and Kiefer, R.W., 1999. *Remote sensing and image interpretation*. John Wiley and Sons, New York, 384 pp.

Martínez-Fernández, J. and Ceballos, A., 2005. Mean soil moisture estimation using temporal stability analysis. *J. Hydrol.*, 312, pp. 28-38.

Pons, X. and Solé-Sugrañés, L., 1994. A Simple Radiometric Correction Model to Improve Automatic Mapping of Vegetation from Multispectral Satellite Data. *Remote Sens. Environ.*, 48, pp. 191-204.

Toutin, T., 2004. Geometric Processing of Remote Sensing Images: Models, Algorithms and Methods (review paper). *Int. J. Remote Sens.*, 25 (10), pp. 1893-1924

Vincent, S. and Pierre, F., 2003. Identifying Main Crop Classes in an irrigated area using High Resolution Image Time Series, IGARSS 2003, pp. 252 - 254.

Wegehenkel, M. and Kersebaum, K.C., 2005. The validation of a modelling system for calculating water balance and catchment discharge using simple techniques based on field data and remote sensing data. *Phys. Chem. Earth*, 30, pp. 171-179.

Willmott, C.J., 1982. Some comments on the evaluation of model performance. *B. Am. Meteorol. Soc.*, 63, pp. 1309−1313.

Wolfe, R.E. Masahiro, N., Fleig, A. J., Kuyper, J. A., Roy, D. P. and Storey, J. C., 2002. Achieving sub-pixel geolocation accuracy in support of MODIS land science. *Remote Sens.* Environ., 83, pp. 31– 49.

ACKNOWLEDGEMENTS

MORE and EBHE Project are funded by the projects REN2003-02956/HID and CGL2008-04047 (Spanish Plan of Science and Technology). This work would not have been possible without aid from projects ESP2006-00643 and ESP2007-65667-C04-04 from the Spanish Ministry of Science and Technology, and AGRHIMOD from the Junta de Castilla y León. The authors also wish to thank the STIG of the University of Salamanca for providing satellite images and other cartographic data and the Regional Government of Castilla y León for the SIGPAC information.

DETECTION OF 25-YEAR LAND-COVER CHANGE IN A CRITICAL WATERSHED IN SOUTHERN PHILIPPINES USING LANDSAT MSS AND ETM+ IMAGES: IMPORTANCE IN WATERSHED REHABILITATION

J. R. Santillan [a, *], M. M. Makinano [b], E. C. Paringit [a]

[a] Research Laboratory for Applied Geodesy and Space Technology, Department of Geodetic Engineering, University of the Philippines (UP), Diliman, Quezon City 1101, Philippines – jrsantillan@up.edu.ph, ecp@engg.upd.edu.ph
[b] College of Engineering and Information Technology, Caraga State University (CSU), Ampayon, Butuan City 8600, Philippines – meriam.makinano@gmail.com

KEY WORDS: Land Cover, Hydrology, Impact Analysis, Modelling, Landsat

ABSTRACT:

We analyzed Landsat MSS and ETM+ images to detect 25-year land-cover change (1976-2001) in the critical Taguibo Watershed in Mindanao Island, Southern Philippines. This watershed has experienced historical modifications of its land-cover due to the presence of logging industries in the 1960s, and continuous deforestation due to illegal logging and slash-and-burn agriculture in the present time. To estimate the impacts of land-cover change on watershed runoff, we utilized the land-cover information derived from the Landsat images to parameterize a GIS-based hydrologic model. The model was then calibrated with field-measured discharge data and used to simulate the responses of the watershed in its year 2001 and year 1976 land-cover conditions. The availability of land-cover information on the most recent state of the watershed from the Landsat ETM+ image made it possible to locate areas for rehabilitation such as barren and logged-over areas. We then created a "rehabilitated" land-cover condition map of the watershed (re-forestation of logged-over areas and agro-forestation of barren areas) and used it to parameterize the model and predict the runoff responses of the watershed. Model results showed that changes in land-cover from 1976 to 2001 were directly related to the significant increase in surface runoff. Runoff predictions showed that a full rehabilitation of the watershed especially in barren and logged-over areas will likely to reduce the generation of huge volume of runoff during rainfall events. The results of this study have demonstrated the usefulness of multi-temporal Landsat images in detecting land-cover change, in identifying areas for rehabilitation, and in evaluating rehabilitation strategies for management of tropical watersheds through its use in hydrologic modeling.

1. INTRODUCTION

Human-induced land-cover changes pose negative impacts to watershed ecosystems. It has been widely recognized that changes such as forest cover reduction through deforestation and conversion for agricultural purposes can alter a watershed's response to rainfall events, that often leads to increased volumes of surface runoff and greatly increase the incidence of flooding (McColl and Aggett, 2007; Cebecauer and Hofierka, 2008). The detection of these changes is crucial to provide information as to what and where the changes have occurred and to analyze these changes in order to formulate proper mitigation measures and rehabilitation strategies.

Remote sensing (RS) techniques have been used extensively to provide accurate and timely information describing the nature and extent of land resources and changes over time. In watershed research and hydrological sciences, RS has played a major role because of its ability to provide spatially continuous data, its potential to provide measurements of hydrological variables not available through traditional techniques, and its ability to provide long term, global-wide data, even for remote and generally inaccessible regions of the Earth (Engman, 1996). It is perhaps for land-cover data derivation that RS has made its largest impact and comes closest to maximize its capabilities in watershed research (Engman, 1995). This has prompted researchers and watershed planners to exploit land-cover information derived from remotely-sensed images in a variety of hydrological modeling studies, most especially in surface runoff predictions (Melesse and Shih, 2002; Bach et al., 2003; Pandey et al., 2008). The addition of Geographic Information System (GIS) technology further enhanced these capabilities and added confidence in the accuracy of modeled watershed conditions, improved the efficiency of the modeling process and increased the estimation capability of hydrologic models (Bhuyan, et al., 2003).

A common approach in integrated RS-GIS-Modeling for event-based watershed runoff predictions usually involves (1) the derivation of land-cover related parameters of the models from remotely-sensed images, (2) the use of GIS to prepare the model and to extract additional parameters, (3) calibration and validation of the model using field measured data to test its efficiency, and then (4) use the model to simulate runoff and use the simulated information to characterize the conditions of the watershed (O'Connell, et al., 2007; Vafeidis, et al., 2007). For land-cover change impact predictions in watersheds, the same approach is generally followed except that the model is run first for an initial land-cover condition, then the land-cover related parameters of the models are altered to reflect the change, and the model is re-run (O'Connell, et al., 2007). The effect of the change is estimated based on the differences between the runoff hydrographs simulated in the initial and "changed" conditions, respectively. Several studies have utilized the RS-GIS-Modeling approach for assessing the impacts of land-cover change to the hydrologic response of

* Corresponding author.

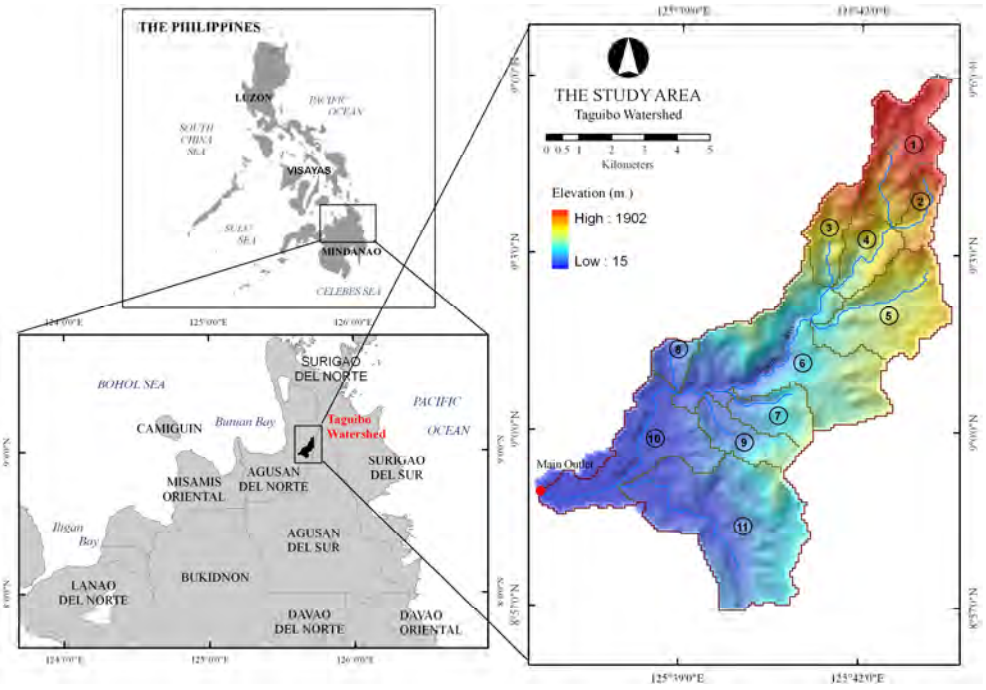


Figure 1. The Taguibo Watershed in Mindanao Island, Philippines

watersheds to rainfall events (e.g., Helmschrot and Flügel, 2002; McColl and Aggett, 2007; Leblanc, et al., 2008). However, a majority of them is focused on modeling the hydrological response of watersheds to future changes in land-cover. Very few studies relate the hydrological responses of watershed to its past and present conditions. In watershed management, this is of paramount importance as the information derived from modeling can be directly related to the changes in land-cover as well as to the overall condition of the modelled watershed. Proper mitigation measures and efficient conservation strategies can then be formulated upon examination of the root causes of watershed problems, and hence, leading to its rehabilitation.

The location and nature of change which has occurred in a watershed can be explicitly recognized using a post-classification comparison approach of land-cover change detection from RS images (Coppin et al., 2004). This approach uses separate classifications of images acquired at different times to produce difference maps from which ''from–to'' change information can be generated (Jensen, 2004). Among the several classifiers available, the Maximum Likelihood Classifier (MLC) has been widely used to classify RS data and successful results of applying this classifier for land-cover mapping have been numerous (e.g., Cherrill et al., 1994; Cingolani, et al., 2004; Hagner and Reese, 2007) despite its limitations due to its assumption of normal distribution of class signatures (Swain and Davis, 1978). Its use has also been effective in a number of post-classification comparison change detection studies (e.g., Helmschrot and Flügel 2002; Muttitanon and Tripathi 2005; Chowdhury 2006; Vagen 2006). While recent studies have indicated the superiority over MLC of newly developed image classifications techniques based on decision trees (DT), neural networks (NN) and support vector machines (SVM) (e.g., Huang et al., 2002; Shuying et al., 2005; Otukei and Blaschke 2010), the advantage of MLC over these classifiers remains to be significant owing to its simplicity and lesser computing time. This is very crucial, especially for rapid land-cover mapping and change detection analysis of numerous multi-temporal images. Moreover, the accuracy of any classifier is affected by the number of training samples and by selecting which bands to use during the classification. As reported by Huang et al. (2002), improved classification accuracies of MLC, DT, NN and SVM can be achieved when training data size is increased and when more bands are included. In the case of Landsat image classification, improvements due to the inclusion of all bands exceeded those due to the use of a better classifier or increased training data size, underlining the need to use as much information as possible in deriving land cover classification from RS images (Huang et al., 2002).

In these contexts, our goal here is to exemplify the importance of land-cover change detection by RS image analysis to provide relevant information on past and present conditions of a watershed. Specifically, we applied post-classification comparison analysis of Maximum Likelihood-classified Landsat MSS and ETM+ images to detect 25-year land-cover change in the critical Taguibo Watershed in Mindanao Island, Southern Philippines. We then relate the changes in land-cover to the responses of the watershed to rainfall events using a GIS-based hydrologic model. The model is also used to test planned rehabilitation measures and strategies to approximate their success or failure in addressing the problems of the Taguibo Watershed

2. THE STUDY AREA

The Taguibo Watershed in the province of Agusan del Norte in the island of Mindanao, Philippines (Figure 1) is a region that has experienced extensive alteration of its land-cover ever since the start of operation of several logging industries with Timber License Agreements (TLAs) in the 1960's until the early 1980s (DENR, 2003). Its forest cover was severely reduced by logging and clear-felling, and the former logged-over areas were opened up to intensive farming, thereby accommodating the influx of farmers who were intent in cultivating semi-temperature high value vegetables. More than 25 years later, a recent report by the Department of Environment and Natural Resources (DENR,

2003) indicated a very serious condition of the watershed: significant increase in runoff volume during rainfall events and extensive sedimentation of rivers and streams due to proliferations of eroded areas in the watershed's landscape. While it has yet to be proven, the DENR asserted that the denudation of the watershed's forest cover and its conversion to grasslands as well as for agricultural purposes are the prime reasons for the occurrence of these problems. The situation was further aggravated by the continuous presence of illegal logging and slash-and-burn farming activities in the upland portions of the watershed (DENR, 2003). These alarming situations have prompted the DENR to come up with rehabilitation efforts such as reforestation of formerly logged areas and agro-forestation in highly eroded landscapes to mitigate the problem of increased runoff generation and high rate of sedimentation. While these efforts to address the negative impacts of land-cover change on hydrological functions are necessary, they can only be fruitful if information on the location and extent of the areas that need rehabilitation is available. Moreover, relevant information that portrays space-time relationships of land-cover to hydrological functions is often required to formulate proper mitigation measures and efficient conservation strategies. All of these can be achieved through analysis of multitemporal RS images and hydrologic modeling.

The Taguibo Watershed has a drainage area of 75.532 km^2. It is composed of plains, steep hills and mountains. According to the Taguibo River Watershed Management Plan (DENR, 2003), majority of the soils in the watershed belongs to hydrologic soil group B (loamy and silty-loamy soils) which indicates medium runoff potential (SCS, 1985). Clayey and shallow soils belonging to hydrologic soil group D (high runoff potential) are generally observed in areas with 50% or more slope. The study area has no distinct dry season; pronounced rainfall occurs from November to January.

3. METHODS

3.1 Landsat image analysis and change detection

Orthorectified Landsat MSS and ETM+ images covering the study area acquired on April 17, 1976 (path 120, row 54) and May 22, 2001 (path 112, row 54), with pixel resolution of 57-m and 28.5-m, respectively, were obtained from the Global Land-cover Facility (GLCF), University of Maryland (http://glcf.umiacs.umd.edu). These images are part of the GLCF GeoCover collection which consists of decadal Landsat data which has been orthorectified and processed to a higher quality standard. Documentations on the orthorectification process can be found in the GLCF GeoCover website at http://glcf.umiacs.umd.edu/research/portal/geocover/.

The images were radiometrically corrected to at-sensor radiance using the standard Landsat calibration formulas and constants. A fast atmospheric correction using dark-object subtraction (Schowengerdt, 1997) was also implemented. Normalised Difference Vegetation Index (NDVI) images were also computed from the radiometrically and atmospherically-corrected images. Only the portions of the images covering the study area were subjected to image analysis. Six (6) land-cover classes were identified from the images through visual interpretations with the aid of existing land-cover and topographic maps published by the DENR as references. These include barren areas, built-up areas, forest, grassland, mixed vegetation (combination of forest, tree plantation, shrub land and grassland) and water bodies. In this study, barren areas are defined as those portions of the watershed with exposed soil and in which less than half of an areal unit has vegetation or other cover while built-up areas are those portions of intensive human use with much of the land covered by structures. The forest class is defined as parcels of land having a tree-crown areal density of 10% or more and are stacked with trees capable of producing timber or other wood products. Grasslands are those portions where the natural vegetation is predominantly grasses and/or grass-like plants.

Built-up areas were only detected on the 2001 Landsat ETM+ image. We assumed that built-up areas in the 1976, although present, were limited in extent so that they were not visible in the Landsat MSS images primarily because of the sensor's low spatial resolution. Representative samples of each class were collected from the images for supervised image classification. The training samples were collected in such a way that the assumption of normal distribution of the MLC is satisfied and that the separability of the classes (computed using the Jeffries-Matusita Distance) (Richards and Jia, 1997) is ≥ 1.7. Another independent set of samples were likewise collected for accuracy assessment. A minimum number of 30 ground truth pixels were randomly chosen for each class, following the guidelines of Van Genderen et al. (1978) to obtain a reliable estimate of classification accuracy of at least 90%.

The MLC was used to classify the Landsat images with the inclusion of the NDVI. The accuracy of each classified images were independently assessed. Four measures were used to assess the accuracy of the classified images namely, the overall classification accuracy, kappa statistic, producer's accuracy (PA) and user's accuracy (UA). Initial trials were done to classify the input images using the Minimum Distance, Mahalanobis Distance and Parallepiped classifiers. However, the accuracies of each classified image using these classifiers were significantly lower (<90%) than those of the MLC-classified images. The two resulting land-cover maps were then subjected to post-classification comparison change detection analysis to examine the location, extent and distribution of land-cover change in the study area. The 2001 land-cover map was first re-sampled to 57-m resolution prior to change detection.

3.2 Hydrologic modelling

Hydrologic modeling was performed using the Soil Conservation Service-Curve Number (SCS-CN) model (SCS, 1985). The SCS-CN model, also called the runoff curve number method, for the estimation of direct runoff from storm rainfall is a well established method in hydrologic engineering and environmental impact analyses and has been very popular because of its convenience, simplicity, authoritative origins, and its responsiveness to four readily grasped watershed properties: soil type, land-use/land-cover and treatment, surface condition, and antecedent moisture condition (Ponce and Hawkins, 1996). The popular form of the SCS-CN model is:

$$Q = \frac{(P - I_a)^2}{P - I_a + S} \quad for\ I_a \leq P,\ otherwise\ Q = 0 \tag{1}$$

$$I_a = \lambda S \tag{2}$$

$$S = \frac{25400}{CN} - 254 \tag{3}$$

where P is total rainfall, I_a is initial abstraction, Q is direct runoff, S is potential maximum retention which can range (0, ∞), and λ is initial abstraction coefficient or ratio. All variables are in millimeters (mm) except for λ which is unitless. The initial abstraction I_a includes short-term losses due to evaporation, interception, surface detention, and infiltration and its ratio to S describes λ which depends on climatic conditions and can range (0, ∞). The SCS has adopted a standard value of 0.2 for the initial abstraction ratio (SCS, 1985) but this can be estimated through calibration with field measured hydrologic data. The potential maximum retention S characterizes the watershed's potential for abstracting and retaining storm moisture, and therefore, its direct runoff potential (Ponce and Hawkins, 1996). S is directly related to land-cover and soil infiltration through the parameter *CN* or "curve number", a non-dimensional quantity varying in the range (0–100) and depends on the antecedent moisture condition of the watershed. Higher *CN* values indicate high runoff potential. For normal antecedent moisture conditions (AMCII, 5-day antecedent rainfall (AR) is 0.5 – 1.1 inches), the curve number values for land-cover types and soil textures (hydrologic soil groups B and D) prevalent in the study area are shown in Table 2. The AMCII CN values can be converted to AMCI (dry condition, AR<0.5 inches) and AMCIII (wet condition, AR>1.1 inches) using the formula of Chow, et al. (1988).

Land-cover	AMCII Curve Number (*CN*)	
	Soil Group B	Soil Group D
Barren areas	86	94
Built-up areas	74	86
Forest	55	77
Grassland	61	80
Mixed Vegetation	58	79
Water	98	98

Table 2. AMCII CN values for different land-cover types under hydrologic soil groups B and D which are prevalent in the study area. (Source: SCS, 1985)

The SCS-CN model was implemented using the Hydrologic Engineering Center-Hydrological Modeling System or HEC-HMS (USACE, 2000). The SCS-CN model was co-implemented with the Clark Unit Hydrograph method (for sub-watershed routing of runoff), the Exponential Baseflow Recession model, and the Muskingum-Cunge model for channel routing. A thorough discussion of these three additional models can be found in Chow et al. (1988). Model parameterizations were done using HEC-GeoHMS (USACE, 2003), the ArcView GIS-based pre-processor of HEC-HMS. HEC-GeoHMS was used to delineate 11 sub-watershed boundaries and reproduce topologically-correct stream network through a series of steps collectively known as terrain pre-processing, by utilizing the surface topography information from a Shuttle Radar Topography Mission (SRTM) DEM as the origin of the boundaries and stream network. Average *CN* values for each sub-watershed were computed based on the 2001 and 1976 land-cover maps. Sub-watershed time of concentration and storage coefficient parameters of the Clark Unit Hydrograph model as well as initial values of the recession constant in each sub-watershed were first assumed but these values were later optimized during the calibration stage. Muskingum-Cunge model parameter values were obtained from river profile and cross-section surveys.

The HEC-HMS model was calibrated using rainfall events recorded at the inner portion of the watershed, and discharge hydrographs measured at the main outlet for the June 25-27, 2007 period. Records of 5-day accumulated rainfall depths before the simulation showed an AR>1.1 inches, indicating AMCIII. Hence, the AMCII values were transformed to AMCIII using Chow, et al. (1988)'s formula. Model calibration was done to fine-tune the λ parameter of the SCS-CN model, and the time-related parameters of the Exponential Baseflow Recession model and Clark Unit Hydrograph model, which were initially assumed. The absence of sources of land-cover information for the state of the watershed when the calibration data were collected prompted us to parameterize the model using the 2001 land-cover map. During this period, available satellite images were all covered with clouds. We assumed that no significant change in land-cover had occurred from 2001-2007. The Nash-Sutcliffe (1970) Coefficient of Model Efficiency, E, was used to evaluate the performance of the hydrologic model during calibration. E ranges between -∞ and 1.0 (1 included) with E = 1 being the optimal value. Values between 0.0 and 1.0 are generally viewed as accepted levels of performance while values ≤ 0.0 indicates that the mean observed value is a better predictor than the simulated value, which indicates unacceptable model performance.

3.3 Runoff predictions in three land-cover conditions

The calibrated hydrologic model was then used to simulate surface runoff in the 11 sub-watersheds under three land-cover conditions namely, 2001, 1976 and a "rehabilitated" condition. The latter was derived from the analysis of the 2001 image, where areas in urgent need of rehabilitation were identified. This includes areas classified as grassland and barren. In the "rehabilitated" land-cover map, grassland areas were re-classified as "forest" while barren areas were converted to "agro-forested areas" which is composed of mixed vegetation. This is in accordance to the rehabilitation strategy planned by the DENR. In using the calibrated hydrologic model for predicting the impacts of land-cover change, as emphasized by the use of the three land-cover condition scenarios, only the *CN* parameter of the model that has a direct relationship with land-cover was altered. The same rainfall events used previously for model calibration were utilized again in the simulations. The results of the simulations were then analyzed (1) to determine the runoff responses of the watershed in 3 land-cover conditions, (2) to identify how different are these responses from each other, and (3) to verify if rehabilitation strategies could help in the reduction of runoff in the watershed under the assumption that the same rainfall events will take place.

4. RESULTS AND DISCUSSION

4.1 Land-cover change in the Taguibo Watershed

The land-cover maps of the study area for 1976 and 2001 derived from Maximum Likelihood-classified Landsat images are shown in Figure 3 (a and b). The 1976 land-cover map has an overall classification accuracy of 96.06% and kappa statistic of 0.95 while the 2001 land-cover map obtained 96.79% accuracy and kappa statistic of 0.96. Producer's and User's Accuracy for each land-cover type are listed in Table 4. It can be observed that the two land-cover maps are above satisfactory because of more than 90% Producer's and User's Accuracy for each land-cover class.

Comparing the two maps, we were able to determine changes in land-cover from 1976-2001 with respect to the total area of the

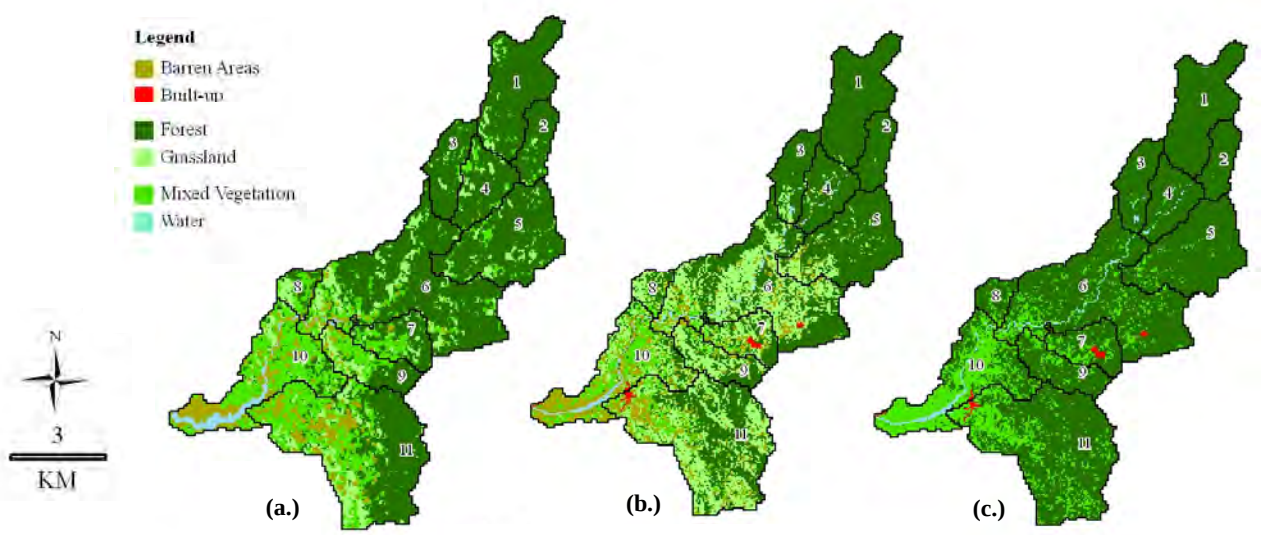


Figure 3. Land-cover maps of the Taguibo Watershed derived from the analysis of Landsat images: *(a.) 1976, (b.) 2001 and (c.) rehabilitated*. Numbers indicate sub-watersheds.

watershed (Table 5). Interestingly, the analysis showed a 6.52% reduction in forest cover, a 13.69% reduction in mixed vegetation, a 4.46% increase in barren areas and 15.54% increase in grassland in the study area in the span of 25 years. The 4.46% increase in barren areas maybe attributed to more recent human-induced alterations of the watershed such as increase in agricultural areas, forest denudation due to illegal logging and slash-and-burn farming and harvesting of planted trees (DENR, 2003). A portion of the 6.52% reduction in forest cover maybe also due to these mentioned activities. On the other hand, the reduction in mixed vegetation cover and increased in grassland areas may be the result of the historical modification of the watershed landscape by logging industries and the influx of farmers who were intent to cultivate the logged-over areas by planting high-valued vegetables and rice crops. When the potential for agricultural productivity of these areas have lessened through time, these were left over for grasses to grow (DENR, 2003). A very good basis of this is the 15.54% increase in grassland areas. As shown in the next sections, these changes in the watershed's land-cover definitely will have an effect on its hydrologic functions, especially on its runoff response to rainfall events.

Land-cover classes	1976 Land-cover Map		2001 Land-cover Map	
	PA	UA	PA	UA
Barren areas	92.50	94.87	98.15	92.98
Built-up areas	100.00	92.75	100.00	100.00
Forest	98.39	98.39	98.96	95.00
Grassland	93.10	98.18	94.51	100.00
Mixed Vegetation	93.33	96.55	100.00	100.00
Water	92.50	94.87	91.67	100.00

Table 4. Producer's (PA) and User's Accuracy (UA) of the land-cover maps (in %).

Land-cover classes	1976 Area (km^2)	2001 Area (km^2)	% Change from 1976 with respect to total watershed area
Barren areas	5.201	8.569	+4.46
Built-up areas	-	0.300	+0.40
Forest	46.287	41.366	-6.52
Grassland	7.271	19.008	+15.54
Mixed Vegetation	15.703	5.359	-13.69
Water	1.070	0.930	-0.19

Table 5. 1976-2001 land-cover change statistics.

4.2 Hydrologic model predictions of runoff

Figure 6 shows the results of the calibration of the hydrologic model with field measured data for the June 25-27, 2007 period. The computed *E* value is 0.92 indicating a highly acceptable performance. However, there are portions of the simulated hydrograph that overestimate the outflow and underestimate the peak discharge. The average residual was computed as 2.95 m^3/s. Plausible explanations for these slight differences in the simulated and measured hydrographs are the fact that the land-cover information used to parameterise during model calibration may be different to the actual land-cover of the study area when the field data were collected. Nevertheless, as the computed *E* value is very close to 1, the model could be used with modest efficiency for runoff predictions under different land-cover conditions of the watershed.

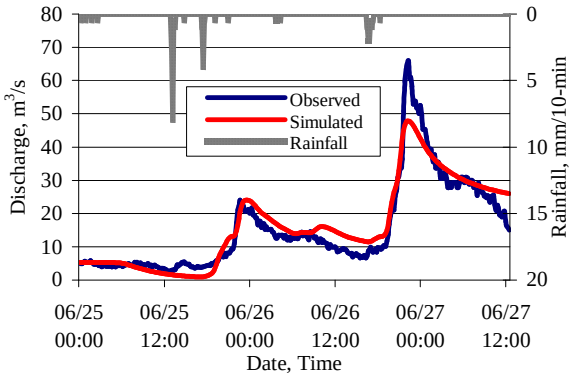


Figure 6. Hydrologic model calibration result.

4.3 Runoff predictions in 3 land-cover conditions

Figure 3c shows the "rehabilitated" land-cover map of the watershed. In this map, the watershed is in an ideal condition where barren areas and grasslands detected from the 2001

Landsat ETM+ image as consequences of anthropogenic disturbances, were rehabilitated through their conversion to mixed vegetation and reforestation, respectively.

Model predicted accumulated runoff volume at each outlet of the 11 sub-watersheds is shown in Figure 7. It can be observed that there were minimal differences in the accumulated runoff volumes in sub-watersheds 1, 2, 3 and 4 for the three land-cover conditions. This means that these sub-watersheds experienced minimal changes in land-cover. The graph also illustrated the high runoff potential of these particular sub-watersheds. Although the majority of land-cover in these areas is forest, the runoff generated during rainfall events is high. This demonstrates the effects of steep slopes and the shallowness of the soil in these areas that give minimal span for the rainwater to infiltrate the ground.

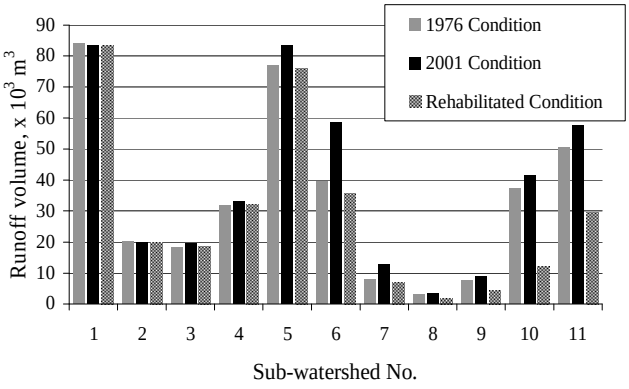


Figure 7. Model predicted accumulated runoff volume for the 11 sub-watersheds from June 25-27, 2007 under 3 land-cover conditions.

Pronounced variation in runoff volumes were observed for the remaining watersheds in the 1976 and 2001 land-cover conditions, most especially in sub-watersheds 5, 6, 7, 9, 10 and 11. It can be stated that changes in major land-cover types in these areas, specifically the increase of barren areas and grasslands and the decrease in forest and mixed vegetation covers (Table 8) have directly affected the hydrologic response of the watershed to rainfall events – rainfall interception and infiltration have been affected such that huge volumes of surface runoff are generated. In terms of total surface runoff accumulated at the main outlet of the watershed (Table 9), model predictions showed that accumulated runoff volume in 1976 were 10.62% lesser than in 2001. Rehabilitation of the sub-watersheds through planting of mixed vegetation and reforestation was found effective. We computed that it reduced the accumulated runoff volume in 2001 by 23.85%. These results provide quantitative estimations that rehabilitation strategies proposed by the DENR, should they be 100% implemented, are most likely to reduce the volume of runoff generated during rainfall events in the Taguibo Watershed.

SW No.	Area, km^2	% change in Barren Areas	% change in Forest	% change in Grassland	% change in Mixed Vegetation
5	8.748	+3.80	-3.68	+5.91	-6.20
6	16.483	+7.27	-26.95	+27.05	-8.96
7	3.224	+12.52	-18.22	+25.93	-23.51
9	3.459	+2.71	+1.91	+24.05	-27.82
10	9.056	+9.01	+3.35%	+13.64	-21.12
11	16.540	+2.39	-5.66%	+23.84	-21.20

Table 8. Major land-cover change from 1976-2001 in sub-watersheds (SW) 5, 6, 7, 9, 10 and 11. Percentage of change is computed with respect to the area of the sub-watershed.

The results of the hydrologic model simulations indicate that significant increase in runoff volumes during rainfall events can be attributed to the reduction in forest and mixed vegetation cover due to their conversions to grasslands and barren areas. These results are consistent with those of Costa et al. (2003) and Siriwaderna et al. (2006).

Land-cover condition	Accumulated watershed runoff volume, $x10^3$ m^3	% Difference from the 2001 condition
1976	376.771	- 10.62%
2001	421.540	
Rehabilitated	320.996	-23.85%

Table 9. Accumulated runoff volumes in 3 land-cover conditions (total for 11 sub-watersheds) for the June 25-27, 2007 period.

5. CONCLUSIONS

We have presented an analysis of 25-year land-cover change in the critical Taguibo Watershed in Mindanao, Philippines using post-classification comparison analysis of Maximum Likelihood-classified Landsat images. We expanded our analysis by incorporating the detected changes in land-cover to a GIS-based hydrologic model. This allowed us to better understand the impacts of the land-cover change to the increase in surface runoff during rainfall events in the Taguibo Watershed. The Landsat image analysis also provided us a very quick identification of areas that need rehabilitation. Using the hydrologic model, we tested planned rehabilitation strategies that were aimed to reduce surface runoff, and we were able to express the effectiveness of these strategies. One of the most important results of this study is that we were able to establish the direct relationship between forest and mixed vegetation cover reduction to increases in surface runoff.

In conclusion, this study has demonstrated the usefulness of multi-temporal Landsat images in detecting land-cover change, in identifying areas for rehabilitation, and in evaluating rehabilitation strategies for management of tropical watersheds through its use in hydrologic modeling. Although the methods used in this study was applied in a relatively small watershed, its applicability to large watersheds and river basins is also possible as long as there are available Landsat images to derive land-cover information needed for detecting and locating the changes, and for hydrologic modeling. With the availability over the internet of Landsat images acquired since 1972, the methods employed in this study can be readily applied for watershed land-cover change monitoring, management and rehabilitation. Moreover, this study made use of the MLC in Landsat image classification. While the land-cover maps derived from the classifications are highly accurate, it is mainly due to the satisfaction of the assumptions of the MLC for class signatures to be normally distributed, and to the high degree of separability of the class signatures. In some cases, the number of training samples to obtain class signatures is limited and/or may not have normal distributions, which restricts the MLC to get the ideal result. The use of other classifiers such as DT, NN and SVM can solve this problem but at the cost of increase in computation time.

REFERENCES

Bach, H., Braun, M., and Mauser, W., 2003. Use of remote sensing for hydrological catchments. *Hydrology and Earth System Sciences*, 7(6), pp. 862-876.

Bhuyan, S.J., Koelliker, J.K., Marzen, L.J., and Harrington Jr., J.A., 2003. An integrated approach for water quality assessment of a Kansas watershed. *Environmental Modelling & Software*, 18, pp. 473-484.

Cebecauer, T. and Hofierka, J., 2008. The consequences of land-cover changes on soil erosion distribution in Slovakia. *Geomorphology*, 98(3-4), pp. 187-198.

Cherrill, A.J., Lane, A., and Fuller, R.M., 1994. The use of classified Landsat 5 Thematic Mapper imagery in the characterization of landscape composition: a case study in Northern England. *Journal of Environmental Management*, 40(4), pp. 357-377.

Chow, V.T., Maidment, D.R., and Mays, L.W., 1988. *Applied Hydrology*. McGraw-Hill: New York, USA.

Chowdhury, R., 2006. Landscape change in the Calakmul Biosphere Reserve, Mexico: Modeling the driving forces of smallholder deforestation in land parcels. *Applied Geography*, 26(2), pp. 129–152.

Cingolani, A.M., Renison, D., Zak, M.R., and Cabido, M.R., 2004. Mapping vegetation in a heterogeneous mountain rangeland using Landsat data: an alternative method to define and classify land-cover units. *Remote Sensing of Environment*, 92(1), pp. 84-97.

Coppin, P., I. Jonckheere, K. Nackaerts, and Muys, B., 2004. Digital change detection methods in ecosystem monitoring: a review. *International Journal of Remote Sensing*, 25, no. 9, pp. 1565–1596.

Costa, M.H., Botta, A., and Cardille, J.A., 2003. Effects of large-scale changes in land cover on the discharge of the Tocantins River, Southeastern Amazonia. *Journal of Hydrology*, 283(1-4), pp. 206-217.

DENR, 2003. *Taguibo River Watershed Management Plan*. Department of Environment and Natural Resources, Caraga Region XIII, Butuan City, Agusan del Norte, Philippines.

Engman, E.T., 1995. The use of remote sensing data in watershed research. *Journal of Soil and Water Conservation*, 50(5), pp. 438-440.

Engman, E.T., 1996. Remote sensing applications to hydrology: future impact. *Hydrological Sciences*, 41(4), pp. 637-647.

Hagner, O., and Reese, H., 2007. A method for calibrated maximum likelihood classification of forest types. *Remote Sensing of Environment*, 110(4), pp. 438-444.

Helmschrot, J. and Flügel, W.A., 2002. Land use characterization and change detection analysis for hydrological model parameterization of large scale afforested areas using remote sensing. *Physics and Chemistry of the Earth*, 27(9-10), pp. 711-718.

Huang, C., Davis, L.S., and Townshend, J.R.G., 2002. An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23(4), pp. 725–749.

Jensen, J.R., 2004. *Introductory Digital Image Processing: A Remote Sensing Perspective*. New Jersey: Prentice-Hall.

Leblanc, M.J., Favreau, G., Massuel, S., Tweed, S.O., Loireau, M., and Cappelaere, B., 2008. Land clearance and hydrological change in the Sahel: SW Niger. *Global and Planetary Change*, 61(3-4), pp. 135-150.

McColl, C. and Aggett, G., 2007. Land-use forecasting and hydrologic model integration for improved land-use decision support. *Journal of Environment Management*, 84(4), pp. 494-512.

Melesse, A.M., and Shih, S.F., 2002. Spatially distributed storm runoff depth estimation using Landsat images and GIS. *Computers and Electronics in Agriculture*, 37, pp. 173-183.

Muttitanon, W., and Tripathi, N.K., 2005. Land use/land cover changes in the coastal zone of Ban Don Bay, Thailand using Landsat 5 TM data. International Journal of Remote Sensing, 26(11), pp. 2311–2323.

Nash, J., and Sutcliffe, J., 1970. River flow forecasting through conceptual models, I - A discussion of principles. *Journal of Hydrology*, 10, pp. 282-290.

O'Connell, E., Ewen, J., O'Donnell, G., and Quinn, P., 2007. Is there a link between agriculture land-use management and flooding? *Hydrology & Earth System Sciences*, 11(1), pp. 96-107.

Otukei, J., and Blaschke, T., 2010. Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *International Journal of Applied Earth Observation and Geoinformation*, 12(Supplement 1), S27-S31.

Pandey, A., Chowdary V.M., Mal, B.C., and Billib M., 2008. Runoff and sediment yield modelling from a small agricultural watershed in India using the WEPP model. *Journal of Hydrology*, 348 (3-4), pp. 305-319.

Ponce, V.M., Hawkins, R.H., 1996. Runoff curve number: has it reached maturity? *Journal of Hydrologic Engineering*, 1, pp. 11-19.

Richards, J.A., and Jia, X., 1997. *Remote Sensing Digital Image Analysis, An Introduction*, Third Edition, Springer-Verlag.

Schowengerdt, R.A., 1997. *Remote Sensing: Models and Methods for Image Processing*, Second Edition, Academic Press.

SCS, 1985. *National Engineering Handbook, Section 4: Hydrology*. Soil Conservation Service, US Department of Agriculture, Washington D.C.

Shuying, J., Deren, L., and Jingwen, W., 2005. A comparison of support vector machine with maximum likelihood classification algorithms on texture features. *Proceedings of IEEE*

International Geoscience and Remote Sensing Symposium, pp. 3717-3720.

Siriwaderna, L., Finlayson, B.L., McMahon, T.A., 2006. The impact of land use change on catchment hydrology in large catchments: The Comet River, Central Queensland, Australia. *Journal of Hydrology*, 326, pp. 199-214.

Swain, P.H. & Davis, S.M., 1978. *Remote sensing: the quantitative approach*, McGraw-Hill New York.

USACE, 2000. *Hydrologic Modeling System HEC-HMS Technical Reference Manual*. United States Army Corps of Engineers-Hydrologic Engineering Center: Davis, California.

USACE, 2003. *Geospatial Hydrologic Modelling Extension HEC-GeoHMS User's Manual, Version 1.1*. United States Army Corps of Engineers-Hydrologic Engineering Center: Davis, California.

Vafeidis, A.T., Drake, N.A., and Wainwright, J., 2007. A proposed method for modeling the hydrologic response of catchments to burning with the use of remote sensing and GIS. *Catena*, 70(3), pp. 396-409.

Vagen, T.G., 2006. Remote sensing of complex land use change trajectories—a case study from the highlands of Madagascar. *Agriculture, Ecosystems and Environment*, 115(1-4), pp. 219–228.

Van Genderen, J.L., Lock, B.F., and Vass, P.A., 1978. Remote sensing: statistical testing of thematic map accuracy. *Remote Sensing of Environment*, 7, pp. 3-14.

ACKNOWLEDGEMENTS

Funding for this research was provided by the Philippine Council for Advanced Science and Technology Research and Development of the Department of Science and Technology (PCASTRD-DOST) through a local graduate scholarship grant to the first author. Engrs. Alexander S. Caparas, Jessie Linn P. Ablao (of UP Diliman) and Michelle V. Japitana (CSU) assisted us during the field surveys. The Field Operations Center of the Philippine Atmospheric Geophysical and Astronomical Services Administration (PAGASA) provided the tipping bucket rain gauge used in this study. The authors would like to thank two anonymous reviewers for their helpful comments and suggestions.

REMOTE SENSING AND GIS IN INFLOW ESTIMATION: THE MAGAT RESERVOIR, PHILIPPINES EXPERIENCE

C.J.S. Sarmiento [a], R.J.V. Ayson [a], R.M. Gonzalez [a, *], P.P.M. Castro [b]

[a] Dept. of Geodetic Engineering, University of the Philippines, Diliman, Quezon City 1101 Philippines – (cssarmiento, rvayson, rmgonzalez)@up.edu.ph
[b] Institute of Civil Engineering, University of the Philippines, Diliman, Quezon City 1101 Philippines – ppmcastro@netscape.net

KEY WORDS: Hydrology, Management, Simulation, GIS, Decision Support, Experimental

ABSTRACT:

In managing a multipurpose dam, knowledge of inflow is essential in planning and scheduling discharges for optimal power production and irrigation supply, and flood control. Utilization of satellite imagery improves inflow estimates provided by digital spatial data instead of those from calculations on drawn maps; the former yields measurements over an area instead of extrapolations from point measurements. Using remote sensing data, GIS techniques, and programming in Java®, an Inflow Monitoring from Basin Assessment Calculations (IMBAC) system was developed to estimate inflow in the Magat watershed; its dam is one of the largest multipurpose dams in Southeast Asia. Magat's 117-km^2 reservoir stores water to irrigate roughly 850 km^2 of farmland and its 360-MW hydro-power plant contributes electricity for Luzon, the Philippines' largest island. The reservoir and dam facilities are jointly managed by the National Irrigation Administration and the SN Aboitiz Power Incorporated; but authorization of discharges during extreme weather conditions is with the country's meteorological agency, the PAGASA. With the complex nature of Magat Dam's multi-stakeholder management involving public and private entities with different discharge motivations, a vital decision support system that concerns inflow estimation is paramount.
This paper presents the results of the developed methodology, IMBAC, to estimate inflow using remote sensing data as an alternative to the water-level approach that is currently being used. IMBAC simulations achieved results which capture the behavior of the Magat watershed response. With more field information to further calibrate the approach, it can be used to build scenarios and simulate inflow estimates under varying watershed conditions.

1. INTRODUCTION

Dams are structures built to create a water reservoir, a hydraulic head and a water surface (Vischer & Hager, 1998). Reservoir operation involves water allocation planning, intake and storage, and discharge control. Knowledge of intake or inflow parameters is essential in planning and scheduling dam discharges, measuring and anticipating current and future power production, optimizing its hydropower operations, and preventing floods. Inflow is a measure of the amount of water entering a reservoir (USACE, 2007). The lack of accurate estimation of inflow parameters is one of the main difficulties in real-world reservoir operations management (Fourcade and Quentin, 1994).

Developments in Remote Sensing (RS) have triggered numerous studies on hydrometeorological model creation and calibration due to the ability of RS in providing spatially-distributed input data (Becker & Jiang, 2007; Kongo & Jewitt, 2007; Wu et al., 2007). Utilization of satellite imagery can improve reservoir inflow estimates by providing digital spatial data instead of calculating from drawn maps, and yielding measurements over an entire area instead of extrapolating from point measurements. Patterns from RS imagery can be translated into a deterministic distribution of input data over a wide area on a pixel-by-pixel basis (Brunner et al., 2007). Coupled with the improving capabilities of Geographic Information Systems (GIS) for simulation and data visualization, RS becomes a powerful source of information that can aid decision makers in the management of reservoirs (Kunstmann et al., 2008).

There are a number of efforts to improve inflow estimation using computational methods such as neural computing (Gilmore, 1996; Kote & Jothiprakash, 2008). Researches involving low resolution satellite images for the characterization of watersheds around a reservoir (Gupta, 2002) have been completed and some of them are focused on certain parameters such as land cover, rainfall (Li et al, n.d.), land surface temperature (Rawls et al, n.d.), surface geology (Ticehurst et al, 2006), soil moisture (Vivoni et al, 2006), vegetation (Bormann, 2007), topography and hydraulic roughness (Aberle & Smart, 2003). These researches are used to improve existing decision-making and discharge policies (Avakyan et al, 2002).

Reservoir managers in the Philippines base their inflow estimates on water level information. With the lack of alternative estimation and forecasting abilities the Magat reservoir managers adapt their management policies to the current water level measurements and rainfall statistics. In this paper, we present the integration of satellite-derived information from Remote Sensing (RS), and Geographic Information System (GIS) visualization and simulation capabilities in improving the Magat Dam inflow estimation process.

* Corresponding author

2. THE STUDY AREA

2.1. Spatial characteristics

This case study was conducted in the watershed of Magat River in Luzon, the largest island of the Philippines. It is bound by the latitudes 17°02'08" and 16°06'05" and the longitudes 120°50'00" and 121°30'00". The watershed is 4463.27 km^2 in horizontal area and is administratively divided between the provincial governments of Ifugao, Isabela and Nueva Vizcaya. The reservoir is approximately 350 kilometers from the capital city, Manila.

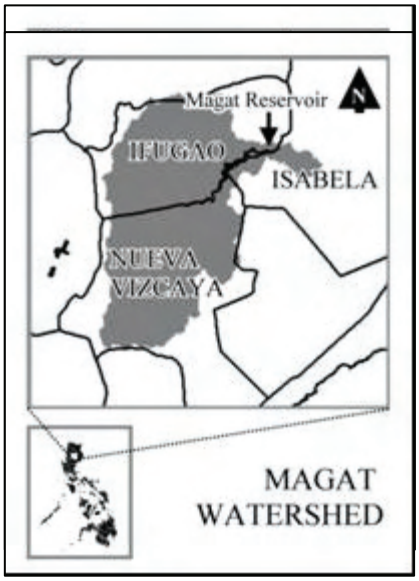


Figure 1. The study area.

2.2. Physical characteristics

The climate of the SW portion of the watershed has two pronounced seasons: wet from May to October and dry from November to April. The rest of the watershed doesn't have pronounced seasons but May to October is relatively wet and the other months are relatively dry (Bato, 2000).

Various forms of clay loam and silt loam soils characterize the whole Magat watershed. Igneous rocks with high silica content (granite and rhyolite) and rocks with low silica content (basalt) as well as scattered sedimentary rocks abound. Four major sets of fault lines run in across the watershed (Palispis, 1979).

2.3. Reservoir structures

Magat Dam is a multipurpose dam which impounds a large reservoir of water from the Magat river. It has a storage capacity of 1.08 billion cubic meters for irrigation to 950 km² of land and 360-MW hydroelectric power generation (Elazegui & Combalicer, 2004). Its flood spillway has a capacity of 32,000 m^3/s.

2.4. Human activities

Agriculture is the most prevalent type of land-use in the watershed. Gold and copper mining interests are also present in Nueva Vizcaya (Elazegui & Combalicer, 2004). Some areas of Ifugao and Isabela are noted for tilapia production and other aquaculture activities.

2.5. Reservoir operation and watershed management

Magat Dam is owned and operated by the National Irrigation Administration. They provide the discharge policy based on a weekly Irrigation Diversion Requirement (IDR). In April of 2007, the operation of the hydroelectric power plant was transferred from the National Power Corporation (NPC) to SN Aboitiz Power (SNAP) after a privatization sale enabled by Republic Act 9136 or the Philippine Electric Power Industry Reform Act (EPIRA) of 2001.

3. MATERIALS AND METHODS

3.1. Landsat imagery

We used the archived 1991, 2002, 2005, 2008 and 2009 8-bit GeoTIFF format images of the study area from Landsat 4, 5 and 7 (L4, L5 and L7). They are designed to capture images over a 185 km swath and gather data at an altitude of 705 km. The study area is within the World Reference System (WRS-2) path 116, rows 48 and 49. The 30m spatial resolution gives sufficient information for the purposes of our study.

3.2. ASTER GDEM

The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) instrument has an along-track stereoscopic capability to acquire stereo image data with a base-to-height ratio of 0.6. B-H ratios between 0.5 and 0.9 are found to be optimal for DEM creation from satellite stereo pairs (Hasegawa et al., 2000). It provides a 1 arcsecond (~30m) resolution, which bodes well with that of the Landsat images. From the ASTER GDEM, the Slope Raster was derived and the Flow Direction Raster was produced using the D8 method.

3.3. TRMM data

TRMM is a joint project of the US National Aeronautics and Space Administration (NASA) and the Japan Aerospace Exploration Agency (JAXA). It was launched from Japan's Tanegashima Space Center on November 27, 1997. The primary purpose of this mission is to observe and estimate rainfall.

3.4. Software

We used ESRI® ArcGIS™ v9.3 (2008) and ITT Industries® ENVI™ v4.3 (2006) as our platforms and Java® for our programming needs.

3.5. Brief overview of the methodology

Let

Qa = reservoir inflow during time, t
Qb = reservoir outflow during time, t
Qr = reservoir storage during time, t

note that

$$Q_b = A\frac{dh}{dt} \quad (1)$$

where A = surface area of reservoir
dh = change in water level height

dt = change in time

We then get the relationship for the storage volume, V.

$$\int (Q_a - Q_b)dt = \int Q_r dt = V \quad (2)$$

If the outflow is zero, we get the relationship:

$$Q_a = Q_b = A\frac{db}{dt} \quad (3)$$

Eq. (3) shows the basic relationship between the water level and the inflow. Through the long experience of the dam managers, they can estimate how much and when the water will enter the reservoir given the rainfall data received from the rain gauges. The volume of the water that enters the reservoir is based on a graph derived from the bathymetry data. A discrepancy due to a time lag variable, the inflow's arrival delay caused by abstractions such as surface roughness (Stephenson & Meadows, 1986), is expected from this procedure because the current method does not account for watershed characteristics quantitatively.

One of the primary uses of remote sensing in this study is the mapping of the Magat watershed's land cover. The eleven classes used were a mix of artificial (cultivated/man-made) and natural features: Cloud, shadow, water, riverwash, fallow field, medium growth field, mature growth field, bare ground, dense forest, sparse forest and built-up.

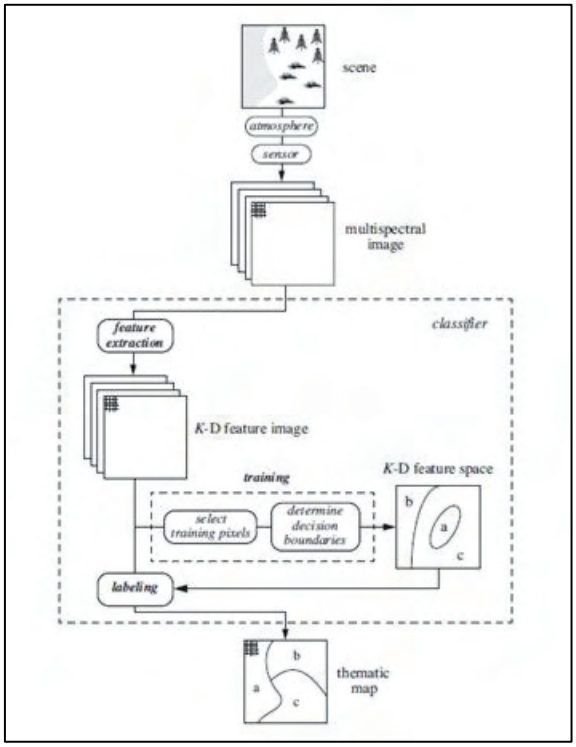


Figure 2. Data flow in a classification process. (Schowengerdt, 2006: p.389)

We tested four types of supervised classifiers (Parallelepiped, Minimum Distance, Mahalanobis Distance and Maximum Likelihood) and two types of unsupervised classifiers (Isodata and K-means). The Maximum Likelihood Classifier produced the highest overall classification accuracy.

Classifier	Overall Accuracy	Kappa Coefficient
Parallelepiped	86.9176%	0.8303
Minimum Distance	86.6824%	0.8260
Mahalanobis Distance	78.1741%	0.7211
Maximum Likelihood	94.7671%	0.9301
Isodata	69.5529%	0.5996
Kmeans	68.8941%	0.5909

Table 4. Overall Accuracy Table

We used the Maximum Likelihood Classifier because it produced the highest overall classification accuracy (Table 4). The classifier assigns the pixels to their corresponding class based on the odds or likelihood that they fit in to that class. The function for each image pixel is calculated by the formula offered by Richards (1999: p.240),

$$g_i(x) = -\ln p(\omega_i) - \frac{1}{2}\ln|\Sigma_i| - \frac{1}{2}(x-m_i)^T\Sigma_i^{-1}(x-m_i) \quad (4)$$

where i = class
x = n-dimensional data (n is the number of bands)
$p(\omega_i)$ = probability that class w_i occurs in the image
$|\sum i|$ = determinant of the covariance matrix of the data in class wi
$\sum i^{-1}$ = its inverse matrix
m_i = mean vector

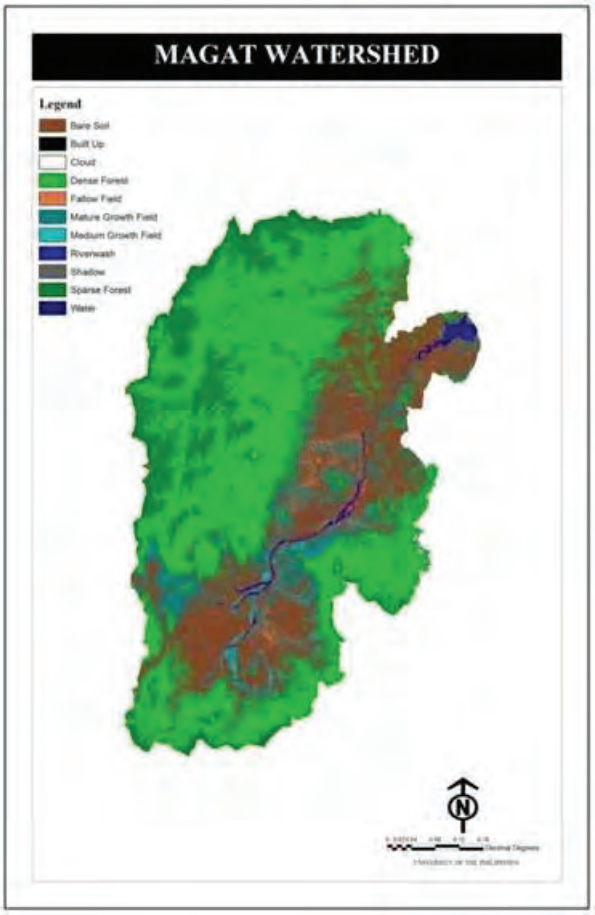


Figure 4. Maximum likelihood classification land cover map of the Magat watershed.

Enhancement and mosaicking of the digital elevation model were done. Thereafter, database building and preparation of the layers in a spatial software environment; Initial extraction of information such as flow direction, accumulation, watersheds and subwatersheds, drainage networks etc., were done using primitive and modified programming scripts.

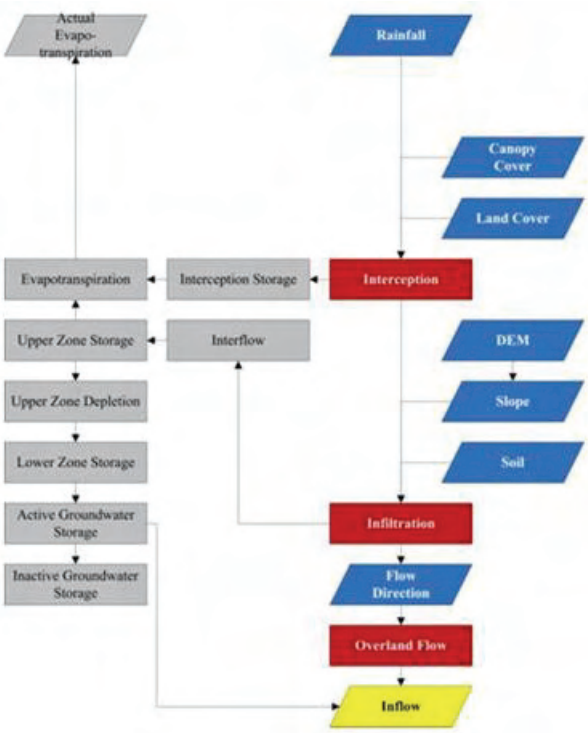


Figure 5. The IMBAC application flowchart

The next stage features the latter part of the hydrologic processing in Java®. The purpose of this application is to create a system that estimates inflow using raster datasets as input. The raster layers are subjected to operators representing hydrologic processes. Figure 5 shows the assembly of the program skeleton and the coding process in its alpha stage.

The initial layers required as input are the rainfall data, canopy cover data, soil cover data and elevation data. The rainfall data from TRMM were combined in a single text file. For the soil and land cover rasters, each class was assigned with representative integers.

From the DEM, the Slope Raster was derived and the Flow Direction Raster was produced using the D8 method. An additional text file representing the digitized reservoir was also included for the application to perform some of its functions like termination.

The Maximum Interception Storage is calculated from the formula modified from von Hoyningen-Huene (1981),

$$S_{MAX}^{C} = 0.935 + 0.498(LAI) - 0.00575(LAI)^2 \quad (5)$$

The Interception Loss for the timestep is therefore calculated using an exponential function by Hedstrom and Polmeroy (1998),

$$I = min\left\{P(1-f_{th,d}) \cdot \left[\frac{S_{max}-S_i(t-1)}{\Delta t}\right] \cdot \left[1-exp\left(-\frac{P(1-f_{th,d})}{S_{max}}\right)\right]\right\} \quad (6)$$

where P = precipitation
S_{max} = maximum interception storage.
S_i (t-1) = interception storage remaining from the previous time step.
$f_{th,d}$ = bypassing fraction.

The infiltration loss model used is proposed by Horton (1939),

$$F_t = f_c t + \frac{(f_0 - f_c)}{k}(1 - e^{kt}) \quad (7)$$

where F_t = infiltration volume at time t.
f_0 = maximum infiltration rate.
f_c = minimum infiltration rate.
k = decay constant.

After calculating for the losses, the resulting virtual run-off matrix is subjected to the flow algorithms. The flow direction raster is one of IMBAC's key-ins. To generate this raster, we used the eight-direction (D8) model described by Jensen and Dominique (1988) wherein a water drop on non-edge pixel can move to that pixel's eight neighbors. It is assumed that the direction of steepest drop is the direction of flow. In order to solve this, we compute for the maximum drop (E_{max})

$$E_{max} = \frac{\Delta z}{D} \quad (8)$$

where Δz = change in the z-value
D = distance between pixel centers

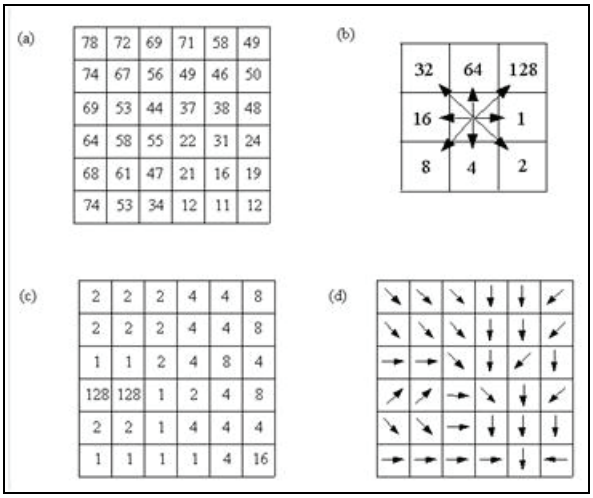


Figure 6. Assignment of flow directions using the D8 model. a. elevations, b. flow direction codes, c. flow direction grid values, d. symbolic representation of flow directions. (NWS, 2008)

The value of the output pixel is specified to indicate the direction of the steepest drop. The following convention (Jensen and Domingue, 1988) is used for the eight valid flow direction representations (E = 1; SE = 2; S = 4; SW = 8; W = 16; NW = 32 ; N = 64; NE = 128). A 5 by 5 expanded neighbourhood is used for instances of more than one pixel having E_{max} values. If the processing pixel is lower than its adjacent pixels, the flow will be undefined – indicating a depression or sink.

4. RESULTS AND DISCUSSION

The graph below shows the comparison of our program's computed inflow and the NIA inflow record for the month of April 2002.

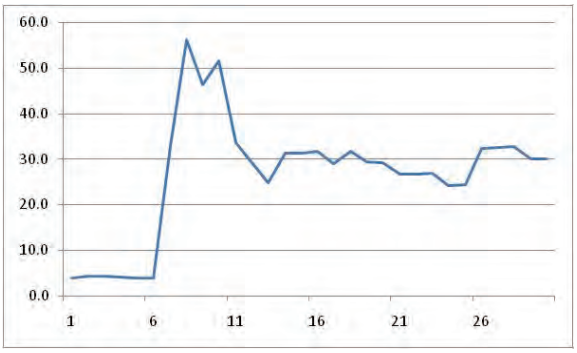


Figure 7. Magat inflow estimates (in m^3/s) using the NIA records.

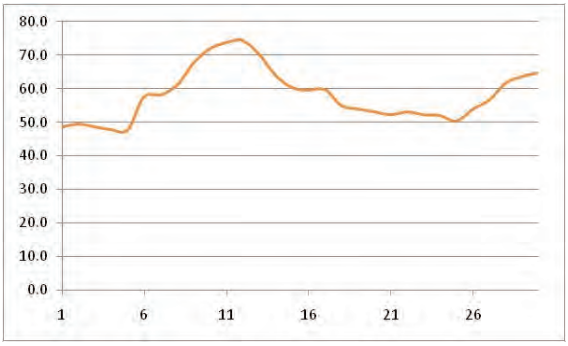


Figure 8. Magat inflow estimates (in m^3/s) using IMBAC.

Differences between the inflow estimates using the current method and the simulated IMBAC inflow estimates can be observed. While IMBAC incorporated the watershed's physical characteristics (canopy cover, slope, soil), meteorological conditions (rainfall), and hydrological processes (infiltration, interception, overland flow), the water level method solely relies on the indicated water level in the dam which corresponds to a tabulated water surface area and its calculated volume. The water level method is purely computational and no modeling or understanding of the watershed hydrology was involved.

It is worth noting however that though there is a discrepancy in the magnitudes of the values produced, the trend of the simulation graphs is similar to the current method graph. The difference during the first week of the simulations and the water level method can be attributed to the discharges made at that time. A discharge cut-off was made on April 6, 2002, and since they base inflow calculations on water level it explains the sudden jump of the graph. The simulations produced by IMBAC produce estimates with a temporal resolution of 30 minutes. This is significantly better than the current method which estimates inflow per day because they capture more detail of the watershed response temporally.

IMBAC simulations achieved results which capture the behavior of the Magat watershed response. With more field information to further calibrate the approach, it can be used to build scenarios and simulate inflow estimates under varying watershed conditions.

5. CONCLUSIONS

Using RS and GIS, we have created a reservoir inflow estimation system, IMBAC, which can be used by the dam managers as an alternative to the current water level-based method. The results produced by IMBAC simulations are promising. The program is proven capable of handling datasets with large extents, thus, it is useful for estimation inflow involving reservoirs with large watersheds. It is capable of preserving the satellite images' pixel characteristics. Each pixel can contain characteristics that are hydrologically relevant; qualitative interpretation of these information is useful in dealing with the scarcity of geographical data at a regional scale. By preserving the pixel characteristics, we can determine the effects of precipitation in one area of the watershed to the inflow estimates--something that the current method cannot do.

In a large study area such as the Magat watershed, logistics play a big role in prioritizing the inclusive activities in this research. The vast size of the study area makes it very difficult to set up streamflow measurement gauges for each sub-watershed. With more time available, more fieldwork to measure streamflow and water velocity should be carried out. Cross-section surveys for each of the drainage segments are also recommended.

An updated source of soil information will also help in refining the inflow estimates. Hydrologic investigations on the study area that focus on infiltration, percolation and groundwater recharge should be done as well.

The densification of the watershed's rain gauge network and an increase in the frequency of recording measurements will also refine inflow estimation efforts. A better system of archiving and securing digital rainfall records is recommended as well. These will also help in assessing the performance of the models used.

The main contribution of this research lies in taking the first steps in realizing the potential of integrating remote sensing and GIS information in reservoir inflow estimation process and Philippine reservoir management in general. We took the first steps in developing a decision support system customized for our country's data situation, economy, policies and multi-stakeholder setup. Studies in improving this framework are currently being undertaken to further refine inflow estimates.

REFERENCES

Aberle, J., & Smart, G.M. 2003. L'influence de la structure de rugosité sur la résistance à des écoulements en fortes pentes. *Journal of Hydraulic Research*. 41(3): 259–269.

Avakyan, A., Litvinov, A. & Riv'er, I. 2002. Sixty Year's Experience in Operating the Rybinsk Reservoir. *Water Resources* 29(1):5-16.

Bato, V. 2000. Determination of major factors affecting the land use/land cover of upper Magat watershed. A preliminary study for the land use/land cover change case study of upper magat watershed. *ACRS 2000*.

Becker, M. & Jiang, Z. 2007. Flux-based contaminant transport in a GIS environment. *Journal of Hydrology* 343:203–210

Bormann, H. 2007. Sensitivity of a soil-vegetation-atmosphere-transfer scheme to input data resolution and data classification. *Journal of Hydrology* 351: 154– 169.

Brunner, P., Hendricks-Franssen, H.J., Kgotlhang, L., Bauer-Gottwein P., & Kinzelbach, W. 2007. How can remote sensing contribute in groundwater modeling? *Hydrogeology Journal* 15:5-18.

Elazegui, D.D. and Combalicer, E.A. 2004. *Realities of the watershed management approach: the magat watershed experience.* PIDS 2004-21.

Fourcade, F., & Quentin, F. 1994. Balancing reservoir management and water conservation: application of hydropower and irrigation. *NATO ASI Series E Applied Sciences* 275:455-464

Gilmore, A. 1996. *A Study of Optimization of Reservoir Operations of the Colorado River*. University of Colorado. USA.

Gupta, H. 2002. Integrated Spatial Data of a Watershed for Planning. *Symposium on Geospatial Theory*.

Hasegawa, H., Matsuo, K., Koarai, M., Watanabe, N., Masaharu, H., Fukushima, Y. (2000). DEM accuracy and the base-to-height ratio of stereo images. *Proceedings of the 33rd ISPRS congress, 4*, 356-360.

Hedström, N.R., & Pomeroy, J.W. 1998. Measurements and modelling of snow interception in the boreal forest. *Journal of Hydrological Processes, 12*, 1611-1625.

Horton, R.E. 1939. Analysis of run-off plot experiments with varying infiltration capacity. *Transactions of American Geophysical Union*, *20*(4), 693-711.

Inflow. *USACE glossary of terms*. 2007. Retrieved (2009, February 6) from http://www.nwk.usace.army.mil/Flood/Glossary.htm

Jensen, S.K., & Domingue, J.O., 1988, Extracting topographic structure for digital elevation data for geographic information systems analysis. Photogrammetric Engineering and Remote Sensing. 54(11): 1593-1600

Kote, A. & Jothiprakash, V. n.d. Reservoir Inflow Prediction Using Time Lagged Recurrent Neural Networks. *Proceedings of the First International Conference on Emerging Trends in Engineering and Technology.*

Kongo, V.M. & Jewitt, G.P.W. 2007. Preliminary investigation of catchment hydrology in response to agricultural water use innovations: a case study of the potshini catchment – south africa. *Physics and Chemistry of the Earth* 31:976–987.

Kunstmann H., Jung, G., Wagner, S., & Clottey H. 2008. Integration of atmospheric sciences and hydrology for the development of decision support systems in sustainable water management. *Physics and Chemistry of the Earth* 33:165–174

Li, R., Spronk, B., & Simons, D. n.d.. *A computer model of rainfall-runoff from a system of multiple watersheds.* Colorado, U.S.A.: Colorado State University

Palispis, J. 1979. *Geology and geochemistry of the oligo-miocene Magat volcanics in the Magat Damsite Ramon, Isabela*. DOST-STII.

Rawls, W., Kustas, W., Schmugge, T., Ritchie, J., Jackson, T., Rango, A., Doraiswamy, P. n.d. *Remote Sensing in Watershed Scale Hydrology*. USDA-ARS Hydrology and Remote Sensing Laboratory. Maryland.

Richards, J.A. 1999. *Remote sensing digital image analysis.* Berlin, Germany: Springer-Verlag.

Schowengerdt, R. 2006. *Remote sensing: models and methods for image processing*. Oxford, U.K.: Elsevier.

Stephenson, D., & Meadows, M.E. 1986. *Kinematic hydrology and modelling*. New York, U.S.A.: Elsevier.

Ticehurst, J., Cresswell, H., McKenzie, N., & Glover, M. 2006. Interpreting soil and topographic properties to conceptualise hillslope hydrology. *Geoderma* 137: 279–292.

Threshold runoff. 2008. Retrieved on February 20, 2010 from http://www.nws.noaa.gov/oh/hrl/gis/data.html

Von Hoyningen-Huene, J. (1981). Die Interzeption Des Niederschlags in landwirtschaftlichen Pflanzenbestanden. (1981). *Arbeitsbericht deutscher verband fur wasserwirtschaft und kulturbau*. Braunschweig, Germany.

Wu, S., Li, J., & Huang, G.H. 2007. Modeling the effects of elevation data resolution on the performance of topography-based watershed runoff simulation. *Environmental Modelling & Software* 22:1250-1260.

ACKNOWLEDGEMENTS

The Philippine Department of Science and Technology-Science Education Institute through the Engineering Research and Development for Technology program for the financial support; USGS-EROS for the satellite images; GES-DISC for the TRMM data; ERSDAC-Japan and NASA-LPDAAC for the GDEM; Mr. Manny Rubio, Mr. Melvyn Eugenio and staff (SN Aboitiz Power), Mr. Pelagio Gamad and staff (National Irrigation Administration-Magat) for the fieldwork and data acquisition assistance

A SUPERVISED SPECTRAL SUBSTRATUM CLASSIFIER TO CLASSIFY IMAGES WITH FUZZY MEMBERSHIPS

Z. Sha[a], Y. Xie[b, *]

[a] Wuhan University, International School of Software, Wuhan, Hubei Province, China - zongyaosha@yahoo.com.cn
[b] Eastern Michigan University, Institute for Geospatial Research & Education, 125 King Hall, Ypsilanti, MI 48197, United States - yxie@emich.edu

KEY WORDS: Classification, Land Cover, Fuzzy Logic, Multispectral, Vegetation, Landsat

ABSTRACT:

Remotely sensed images often display spectral variations over heterogeneous regions in the context of land cover classes (LCCs), which imposes challenges to information extraction from the images. In this paper, an easy-to-apply image classification model, supervised spectral substratum classifier, is proposed. The classifier first builds spectral LCCs (SLCCs) from a training dataset (*TD*). A SLCC comprises the spectral signals of a labeled LCC in *TD* based on the ground truth. This SLCC is further marked as homogeneous or heterogeneous according to the statistical properties of the mean value and the standard deviation of all spectral cases in this SLCC. When this SLCC is marked as heterogeneous, the spectral space of the SLCC will be disaggregated (or clustered) into substrata by applying statistical cluster analysis. A membership function is then defined for each substratum. To classify images, fuzzy membership functions are applied to measure similarities between corresponding spectral substrata and any new to-be-classified cases (pixels). The new cases are classified to the most comparable substrata as determined by the membership functions. As a case study, a vegetation cover classification over a typical grassland in Inner Mongolia from Landsat ETM+ is conducted. The result shows that the proposed classification model obtains an overall accuracy of 79.3% and kappa of 0.76. As comparison, a hybrid fuzzy classifier and a conventional and hard classification of maximum likelihood were applied as references.

1. INTRODUCTION

Remote sensing technology has been proved to be practical and economical means to study land cover changes and to assess natural resources, especially over large areas (Langley et al., 2001; Nordberg and Evertson, 2003). Image classification is widely used to derive useful information from remotely sensed datasets. Various models, or image classifiers, have been developed to extract land cover information from remote sensed images. Image classifiers can be broadly divided into unsupervised ones and supervised ones. Unsupervised approaches are often used in thematic mapping from imagery, and available in most of the image processing and statistical software packages (Langley et al., 2001). For supervised classification, a maximum likelihood (ML) classifier is usually viewed as a classic and most widely used method (Sohn and Rebello, 2002; Xu et al., 2005). More advanced classification models, such as artificial neural network (ANN) and support vector machine (SVM), have been attempted in recent years (Černá and Chytrý, 2005; Cristianini and Shawe-Taylor, 2000; Du and Sun, 2008; Gustavo and Lorenzo, 2009). Fuzzy logic classification, a kind of probability-based classification, also gets good attentions in recent years (Triepke et al., 2008).

To get a better classification result, there have been a few attempts to incorporate different image classification methods. Lo and Choi (2004) developed a hybrid classification method that incorporated the advantages of supervised and unsupervised approaches as well as hard and soft classifications for mapping the land use/cover of the Atlanta metropolitan area using Landsat 7 Enhanced Thematic Mapper Plus (ETM +) data. They applied a supervised fuzzy classification to the mixed pixels, and got a slightly better result than other methods (unsupervised ISODATA, supervised fuzzy, and supervised maximum likelihood classification methods) in terms of land use/cover classification accuracy. Laba et al. (2002) compared the accuracy of a regional-scale thematic map of land cover at taxonomic resolutions (i.e., different classification levels). The study showed that the map produced by the fuzzy-method had an obvious improvement in accuracy at both low and high taxonomic resolutions. In general, fuzzy image classifiers are more suitable for heterogeneous areas, while hard classifications are widely applied in homogeneous areas (Sha et al., 2008).

We propose in this paper an easy-to-apply fuzzy classification model (classifier) to extract land cover classes (LCCs) from remotely sensed images. The classifier first builds spectral LCCs (SLCCs) from a training dataset. A SLCC will be marked as heterogeneous if the statistical properties (mean value and standard deviation) of the cases labeled with this SLCC meet certain criteria. The spectral space of this SLCC will be disaggregated (or clustered) into substrata by applying statistical cluster analysis. Fuzzy membership functions are defined for the substrata based on the training dataset and then applied to measure similarities between the new cases and these spectral substrata and to determine their classifications.

2. METHODOLOGY

Many heterogeneous regions show obvious spectral variations over LCCs in remotely sensed images. Specifically, the cases labeled as a single LCC may demonstrate distince spectral deviations. Under such a condition, the cases labeled as the

* Corresponding author

LCC can be clustered into several sub-groups (substrata) with smaller within-group spectral deviations based on the spectral properties of the LCC labeled cases. This classification process is called a spectral substratum classifier.

Five steps are involved in implementing the proposed classification model. Step 1 creates a classification information system from the training dataset. Step 2 builds spectral substrata (SS) space for each SLCC from the training dataset. Step 3 defines membership functions for assigning new cases. Step 4 evaluates the classifier's performance through a testing dataset. Step 5 classifies new cases by applying the derived classifier.

Step 1: Creating a classification information system from the training dataset

For a given nonempty finite set of cases U={x_t} (t=1,2,……,n) where x_t indicates case t. Each case x_t in U is depicted by a set of attribute variables B_i={b_i} (i=1,2,…,m) and labeled by a class $C_{j.}$ (j=1, 2, …, n), $C_{j.} \in$ C={C_1, C_2, …, C_k}, where b_i has a continuous value domain, C is a priori class label set, and the symbol "." in $C_{j.}$ indicates one of the candidate classes from C. That is to say, x_t = {b_{t1} , b_{t2}, …, b_{tm}, $C_{j.}$} Therefore, U can also be viewed as a matrix M with n rows and m+1 columns,

$$M_{n,m+1}(B,C)=\begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1m} & C_{1\cdot} \\ b_{21} & b_{22} & \cdots & b_{2m} & C_{2\cdot} \\ b_{31} & b_{32} & \cdots & b_{3m} & C_{3\cdot} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ b_{n1} & b_{n2} & \cdots & b_{nm} & C_{n\cdot} \end{pmatrix} \quad (1)$$

whereas the first m columns are called condition variables and the last column is a decision result. $C_{j.}$ denotes different elements from the candidate class set C. Note that different B_i may have the same class label.

Any object in U is uniquely determined by the values of its attributes. In other words, for any object $x_t \in$ U with an attribute set b_i describing x_t, the object can be uniquely classified (labeled) as C_j. This form of U (or M) is usually referred to as a classification information system (CIS). The notion of classification information systems (sometimes called data tables, information tables, attribute-value systems, knowledge representation systems, etc.) provides a convenient tool for the representation of objects in terms of their attribute values. The training dataset is taken to build a classifier and the testing dataset used to test the accuracy of a derived classifier that has the form of CIS.

Step 2: Building SS-Space for each SLCC from the training dataset

Let x_{i,j_z} denotes the j_z^{th} observation within class C_j for variable b_i, with $1 \le j_z \le$ n_j, $1 \le$ n_j $\le$ n, n_j being the number of observations in class C_j, and $\sum_{j=1}^{k} n\_j$ =n. For all cases with such a unique label as C_j in U, calculate the mean values ($\bar{x}_{i,j}$) and the standard deviations (denoted as $\sigma_i(C_j)$) of the observations labeled as C_j for each variable b_i (i=1,2,…, m),

$$\bar{x}_{i,j} = \frac{1}{n\_j}\sum_{j_Z=1}^{n\_j} x_{i,j_Z} \quad (2)$$

$$\sigma_i(C_j) = \sqrt{\frac{(x_{i,j} - \bar{x}_{i,j})^2}{n\_j}} \quad (3)$$

Let $\bar{\sigma}_i$ be the mean value of $\sigma_i(C_j)$ of all C_j (j=1, 2, … k) for variable b_i (i=1, 2, …, m),

$$\overline{\sigma}_i = \frac{1}{k}\sum_{i=1}^{m}\ \sum_{j=1}^{k}\sigma_i(C_j) \quad (4)$$

For m variables (b_i) in CIS, we can get a S.D. vector σ = ($\bar{\sigma}_1$, $\bar{\sigma}_2$, …, $\bar{\sigma}_m$) with dimension m. When $\sigma_i(C_j) \ge \bar{\sigma}_i$ (for variable b_i) happens to a candidate class C_j, there exists a larger deviation of spectral signals among the observations with respect to variable b_i. Therefore, $\sigma_i(C_j) \ge \bar{\sigma}_i$ is used as a judgment to mark "heterogeneous" for these cases. A recursive clustering to these cases labeled as C_j is then performed to determine substrata of the spectral signals until S.D. of each SLCC (subclasses of C_j, denoted as $C_{j.}$ where "·" indicates that the subclasses come from C_j) satisfies $\sigma_i(C_{j.}) < \bar{\sigma}_i$.

A two-step hierarchical clustering analysis is recommended by using the Statistical Package for the Social Sciences (SPSS) (http://www.wright.edu/cats/docs/docroom/spss/), with the original candidate class as priori group and Euclidean distance as the linkage distance measure for variable bi, and the unweighted pair-group centroids as the linkage rule (LR). In addition to the control parameter of $\bar{\sigma}_i$, each subclass has to meet the requirement of a minimum number of cases (*MNC*). When the minimum case requirement is not satisfied, the cases will be merged with its nearest subclass in terms of Euclidean distance.

Accordingly, all the original cases can then be either labeled as C_j if no clustering is needed on the basis of $\bar{\sigma}_i$, or $C_{j\text{-}f}$ (f=1, 2, …, p, where p is the total number of the substrata after the clustered cases are labeled as C_j in the training dataset). In other word, the original k candidate LCC classes can be extended to γ substrata classes where $\gamma = \sum_{j=1}^{k}\sum_{f=1}^{p}1$. For each $C_{j\cdot}$, the combination of $\bar{x}_{i,j\cdot}$ and $\sigma_i(C_{j\cdot})$, similarly calculated by Function (2) and (3), is referred to as a spectral substratum space (SS space) for variable b_i and class $C_{j\cdot}$. A clustering

analysis based on the CIS creates an SS space vector for each b_i (SS_i) with γ elements defined,

$$SS=\{ SS_1, SS_2, \ldots, SS_n\}$$
$$\text{and } SS_i=\{ss_1, ss_2, \ldots, ss_r\}, \quad (5)$$
$$\text{where } ss_i=(\bar{x}_{i,j\cdot}, \sigma_i(C_{j\cdot}))$$

Step 3: Defining membership functions from the training dataset

For any new object (case) that is attributed by b_i (i=1, 2, …, *m*), the classification task is to find out a substratum ($C_{j\cdot}$) that has the most comparable properties to the object. Thus, for any b_i, the following membership function is defined to calculate the similarity measure (*SM*) between the object and the class $C_{j\cdot}$,

$$SM_j = Max\{SM_{j\cdot}\}$$
$$\text{and } SM_{j\cdot} = \begin{cases} Max\ (1-|b_i-\bar{x}_{i,j-f}|/\beta\cdot\sigma_i(C_{j-f})), \text{ after clustering} \\ 1-|b_i-\bar{x}_{i,j}|/\beta\cdot\sigma_i(C_j), \text{ if no clustering is made} \end{cases} \quad (6)$$

where j· represents C_j if no clustering analysis is made, or C_{j-f} (f=1, 2, …, p) if C_j is grouped into p substrata after clustering. β is called overlapping coefficient for $\sigma_i(C_{j\cdot})$. For any C_j, $SM_{j\cdot}$ is the largest value of the similarity measure from the subclasses (i.e., substrata from the C_j). It can be seen that $SM_{j\cdot}$ =1 if $b_i = \bar{x}_{i,j\cdot}$, and $SM_{j\cdot}$ =0 if $|b_i - \bar{x}_{i,j\cdot}| \geq \beta\cdot\sigma_i(C_{j\cdot})$. The class label finally assigned to the object will be the label (C_j) that presents the closest similarity through the Max operation defined in Function 6.

For any new case, the defined membership (given by Function 6) compares the similarity between the new case and a number of known class (C_j or C_{j-f}) determined from the training dataset. Instead of directly assigning the new case to a known class, the function selects the most comparable class, to whichthe new case belong based on the calculated similarity measurement ($SM_{j\cdot}$).

An Example is provided as an illustration:

Suppose, for variable b_1, we have $\bar{x}_{1,1\cdot}$=0.50, $\sigma_i(C_{1\cdot})$ =0.05; $\bar{x}_{2,2\cdot}$ =0.55, $\sigma_1(C_{2\cdot})$ =0.02. A given new case has b_1 =0.58 and β takes 3.0.

Scenario 1: According to Function 6, SM_1=0.47 and SM_2=0.50. This is to say, the new case shows more similarity to C_2 than to C_1. Therefore, the new case is more likely to be C_2 if the variable b_1 is taken as the input for consideration.

Scenario 2: When the deviation of sample cases is considered, the cases labeled as C_1 must be subdivided into two substrata through the clustering analysis, C_{1-1} and C_{1-2}, because of the fact that $\sigma_1(C_{1\cdot})$ =0.05 $\geq \bar{\sigma}_1$ =0.035. After the clustering analysis, we get two SS-spaces for C_{1-1} and C_{1-2} respectively, s_{1-1} = ($\bar{x}_{1,1-1}$ = 0.43, $\sigma_1(C_{1-1})$ = 0.02) and s_{1-2} = ($\bar{x}_{1,1-2}$ = 0.57, $\sigma_1(C_{1-2})$ =0.02).

According to Function 6, SM_1=max {$SM_{1,1-1}$, $SM_{1,1-2}$} = {0, 0.83} = 0.83, and SM_2=0.50. Therefore, the new case is more comparable to C_1 if the variable b_1 is taken as the input for consideration. When there are a set of properties (b_i) to depict an object, a weighted average of SM_j for all considered b_i is used to decide the similarity measure,

$$\overline{SM_j} = \frac{1}{n}\sum_{i=1}^{n}\omega_i \cdot SM_j \quad (7)$$

where $\overline{SM_j}$ is the weighted average SM_j for all considered b_i, and ω_i stands for the weight for variable b_i. We placed equal weights for the 6 image bands in the case study. $\overline{SM_j}$ measures the similarity between any new case and C_j, and is computed to quantify the possibility that the new case belongs to a class C_j. The given object will be assigned to C_j that has the largest $\overline{SM_j}$.

Step 4: Evaluating classifier performance through testing dataset

The classified results derived from remote sensed images should be objectively verified and communicated to users so that they can make informed decisions on whether and how the products can be used. A testing dataset is used to evaluate the degree of 'correctness' of the classified features compared to the actual ones. Though there are a few evaluation methods, we used a confusion (or error) matrix to evaluate the classification result, which describes the goodness of fit between the derived classes and the reference data through using the measures like overall accuracy and kappa coefficient.

The testing dataset has a similar format as that of the training dataset. Any case in the testing dataset has an input vector denoted as {b_1, b_2, …, b_n} and an actual output class label C_j. The built fuzzy classifier from Step 3 takes the input vector and outputs (or assigns) a class label. The classifier outcomes are compared with the actual ones to build the error matrix. The overall classification accuracy and Kappa statistic are calculated to quantify the result (de Leeuw et al. 2006).

Step 5: Classifying new cases by applying the derived classifier

For any new case that needs to be classified, the derived classifier, if a reasonable classification accuracy is achieved, is employed to make classification. To classify a remotely sensed image, each pixel is a new case that is taken as input to the derived classifier and an output class label then is determined. Afterwards, a map covering the image area is sually produced to show visualize the classification result.

3. CASE STUDY

The proposed classifier is applied to classify grassland vegetation from Landsat Enhanced Thematic Mapper (ETM+) in Xilin River Basin, Inner Mongolia, China (Figure 1). The study region covers an area of nearly 10, 000 km^2 and strides two Landsat ETM+ scenes. It is one of the most representative steppe zones in China and the world (Xie et al., 2010). It is known from the long-term observations and field samplings that much of the research region is dominated by heterogeneous plant communities. It is confirmed that hard and pixel-based image classifications were not the right ways to map the vegetation cover in this region (Sha et al., 2008).

A vegetation classification system consisting of 11 vegetation communities is determined based on the plant ecological and biological features (Table 1). Two image scenes of Landsat ETM+ (path 124/row 29 and path 124/row 30) on 14 August, 2004, covering the whole region, are obtained and a series of image preprocessing tasks are performed to produce a qualified image for classifying vegetation cover (Sha et al., 2008). Simultaneous 464 ground samples evenly distributed over the study area were collected in the field with a hand-held global positioning system (GPS) with an accuracy of 15m and geo-registered to the image. The ground samples are divided into two groups: the training dataset and the testing dataset.

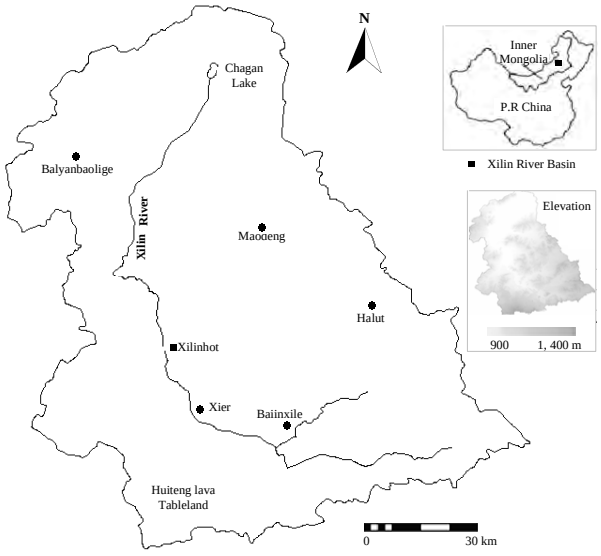


Figure 1. Study area

The training dataset, including 348 samples with six bands (variables) and the class labels (C_1, C_2, …, C_{11}), are taken to build the spectral substratum classifier. The rest 116 samples are used to validate its classification accuracy. All of the samples (training and testing) and the pre-processed image are used to extract spectral data of each ground sample. Spectral data for the samples are analyzed with the six reflective bands of the Landsat ETM+. Bands 1, 2, 3, 4, 5 and 7 of the pre-processed image are analyzed separately to create a SS space vector. The six bands of the pre-processed image are normalized through the following function, respectively,

$$b_i = (b - b_{\mathrm{Min}}) / (b_{\mathrm{Max}} - b_{\mathrm{Min}}) \quad (8)$$

where b_i is the transformed value of any original pixel value (b) for layer i that has maximum and minimum pixel values given by b_{Max} and b_{Min}.

class	Community type (named after dominant species)	Vegetation type
C_1	*Cleistogenes squarrosa*	Typical steppe
C_2	*Stipa grandis*	Typical steppe
C_3	*Achnatherum splendens*	Meadow
C_4	*Stipa krylovii*	Typical steppe
C_5	*Artemisia frigida*	Typical steppe
C_6	*Carex pediformis*	Meadow steppe
C_7	*Carex spp.*	Meadow
C_8	*Caragana microphylla*	Typical steppe
C_9	*Leymus chinensis+Stipa baicalensis*	Meadow steppe
C_{10}	*Leymus chinensis*	Typical steppe
C_{11}	*Salsola collina (Chenopodium glaucum)*	Typical steppe

Table 2. Vegetation classification system

The 'brightness' value of each image band of the samples in the training set was normalized according to Equation 8 to produce a training data matrix of 348 × 7 (six bands + vegetation type). Similarly, all testing samples were processed to form a testing data matrix of 116 × 7. These matrixes are two CISs used to facilitate our analyses. The six variables (bands) along with the class label from the training matrix were analyzed to form a SS space vector {SS_1, SS_2, SS_3, SS_4, SS_5, SS_7}. It was found that the cases labeled with C_1, C_2, C_4, C_5, C_{10}, and C_{11} in the training dataset for most of the bands were marked heterogeneous and thus clustering analysis was applied to form substrata. Take band 1 as an example, a clustering analysis was performed on the cases originally labeled as C_2, C_4, and C_{10} in the training dataset. The results of the clustering analyses were reported Table 3. In other words, the cases in band 1 labeled as C_2, C_4, or C_{10} in the training dataset displayed significant variations, forming seven substrata. Moreover, the spectral variation of the variable within each substratum was decreasing when it was further clustered. As Table 2 revealed, the largest variations occurred with the cases labeled as C_2 since most LCCs in C_2 produced three substrata, $C_{2\text{-}1}$, $C_{2\text{-}2}$, and $C_{2\text{-}3}$ (Table 2).

The constructed classifier was then applied to the testing dataset for an accuracy evaluation. The result of the accuracy test showed that most reference classes with large sizes of cases could be well predicted by the classifier. The overall accuracy of the classifier reaches 79.3% with Kappa valued of 0.76 (Table 3). Considering the accuracy obtained from the classifier, the spectral substratum classifier could be applied to classify the whole image according to Equations 6)and 7 to derive the final vegetation cover map over the study region.

SS_{band}	SS_i list	γ
SS_1	C_1(0.52, 0.03) $C_{2\text{-}1}$(0.18, 0.04) $C_{2\text{-}2}$(0.31, 0.04) $C_{2\text{-}3}$(0.63, 0.04) C_3(0.16, 0.02) $C_{4\text{-}1}$(0.12, 0.02) $C_{4\text{-}2}$(0.67, 0.02) C_5(0.49, 0.04) C_6(0.62, 0.02) C_7(0.55, 0.03) C_8(0.96, 0.02) C_9(0.40, 0.02) $C_{10\text{-}1}$(0.20, 0.04) $C_{10\text{-}2}$(0.50, 0.03) C_{11}(0.52, 0.02)	15
SS_2	$C_{1\text{-}1}$(0.38, 0.03) $C_{1\text{-}2}$(0.64, 0.02) $C_{2\text{-}1}$(0.21, 0.02) $C_{2\text{-}2}$(0.31, 0.03) $C_{2\text{-}3}$(0.53, 0.03) C_3(0.67, 0.02) $C_{4\text{-}1}$(0.37, 0.03) $C_{4\text{-}2}$(0.57, 0.04) $C_{5\text{-}1}$(0.35, 0.02) $C_{5\text{-}2}$(0.73, 0.03) C_6(0.36, 0.03) C_7(0.54, 0.02) C_8(0.70, 0.02) C_9(0.27, 0.02) $C_{10\text{-}1}$(0.20, 0.04) $C_{10\text{-}2}$(0.24, 0.04) C_{11}(0.42, 0.02)	17
SS_3	C_1(0.61, 0.03) $C_{2\text{-}1}$(0.30, 0.02) $C_{2\text{-}2}$(0.60, 0.02) C_3(0.37, 0.02) $C_{4\text{-}1}$(0.24, 0.03) $C_{4\text{-}2}$(0.50, 0.03) C_5(0.46, 0.02) C_6(0.28, 0.02) C_7(0.76, 0.04) C_8(0.20, 0.03) C_9(0.53, 0.02) $C_{10\text{-}1}$(0.15, 0.04) $C_{10\text{-}2}$(0.33, 0.02) $C_{11\text{-}1}$(0.29, 0.03) C11-2(0.34, 0.02)	15
SS_4	C_1(0.32, 0.02) $C_{2\text{-}1}$(0.19, 0.03) $C_{2\text{-}2}$(0.50, 0.02) $C_{2\text{-}3}$(0.69, 0.03) C_3(0.44, 0.03) $C_{4\text{-}1}$(0.37, 0.03) $C_{4\text{-}2}$(0.86, 0.02) $C_{5\text{-}1}$(0.27, 0.02) $C_{5\text{-}2}$(0.60, 0.03) C_6(0.65, 0.02) C_7(0.23, 0.02) C_8(0.36, 0.02) C_9(0.30, 0.02) $C_{10\text{-}1}$(0.28, 0.02) $C_{10\text{-}2}$(0.53, 0.04) C_{11}(0.45, 0.03)	15
SS_5	$C_{1\text{-}1}$(0.28, 0.02) $C_{1\text{-}2}$(0.39, 0.04) $C_{2\text{-}1}$(0.25, 0.04) $C_{2\text{-}2}$(0.44, 0.02) $C_{2\text{-}3}$(0.83, 0.03) C_3(0.49, 0.03) $C_{4\text{-}1}$(0.28, 0.03) $C_{4\text{-}2}$(0.77, 0.03) C_5(0.39, 0.04) C_6(0.30, 0.02) C_7(0.55, 0.02) C_8(0.74, 0.03) C_9(0.68, 0.02) $C_{10\text{-}1}$(0.26, 0.03) $C_{10\text{-}2}$(0.72, 0.03) C_{11}(0.31, 0.02)	16
SS_7	$C_{1\text{-}1}$(0.26, 0.02) $C_{1\text{-}2}$(0.59, 0.03) $C_{2\text{-}1}$(0.52, 0.03) $C_{2\text{-}2}$(0.78, 0.02) C_3(0.45, 0.02) $C_{4\text{-}1}$(0.32, 0.04) $C_{4\text{-}2}$(0.54, 0.03) $C_{5\text{-}1}$(0.23, 0.02) $C_{5\text{-}2}$(0.45, 0.03) C_6(0.35, 0.04) C_7(0.23, 0.03) C_8(0.51, 0.04) C_9(0.23, 0.03) $C_{10\text{-}1}$(0.18, 0.02) $C_{10\text{-}2}$(0.31, 0.03) C_{11}(0.45, 0.03)	15

Table 3. SS space vector derived from the training dataset (*MNC* (minimum number of cases) =5)

Map class	Reference class											Total	User's
	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}		
C_1	5	2		1								8	62.5
C_2	1	23	1	1		1			1	1		29	79.3
C_3			6								1	7	85.7
C_4		1		15								16	93.8
C_5	2				4							6	66.7
C_6						2						2	100
C_7					1		2					3	66.7
C_8								6		1		7	85.7
C_9					1				4			5	80
C_{10}		2		2			1			15		20	75
C_{11}				1		1				1	10	13	76.9
Total	8	28	7	20	6	4	3	6	5	18	11	116	
P*	63	82	86	75	67	50	67	100	80	83	91		

Overall accuracy: 79.3%; kappa=0.76
P* standards for Producer's

Table 4. Error matrix for the proposed classifier

The classification performance was further assessed in comparison with other classification models (Table 5). It was found that the result from the proposed substratum classifier produced a comparable accuracy as the hybrid fuzzy classifier (HFC) did in the same study area (Sha, et al., 2008) and had a much better performance than the conventional supervised classifier (CSC) model. In addition, in terms of the procedures involved, the proposed substratum model was relatively easier to carry out.

Classification method	Percentage classified	Kappa
HFC(Sha et al., 2008)	80.2	0.77
CSC*	69.0	0.63
Spectral Substratum Model	79.3	0.76

*CSC: Conventional supervised classification on the basis of maximum likelihood.

Table 5. Result comparison with other models

4. CONCLUSION AND DISCUSSIONS

The proposed spectral substratum classifier essentially adopts a fuzzy or soft classification strategy. Fuzzy classifiers have been studied for years and proved to produce more accurate classifications compared to the hard methods especially over a hetegeneous environment. Under such a condition,, a pixel in an image may not display an overwhelming similarity to a LCC. Instead, it would be better to say that the pixel is more likely belonging to a LCC.

In the applications of environmental mapping from remote sensed images, two considerations are usually taken into account to develop a new classification system, when the sampled cases in the training dataset show distinct spectral variations even if they belong to the same LCC. First, if the spectral variations are within a reasonable limit, all cases can be treated as a class corresponding to a LCC. Second, if the variations within a spectral LCC (SLCC) are too large, these cases can be split into two or more subclasses (substrata). The derived substrata will be used to replace the original SLCC. Compared to SLCC, these substrata show much smaller within-group spectral variations. In our research, a hierarchical clustering analysis is performed with all the variables in the training dataset as an input vector to derive substrata for the LCC cases. Each substratum has a membership function defined by the statistical properties (mean value and standard deviation) of the cases labeled with this substratum. During the clustering process for variable *i* and class C_j, two parameters are examined to control its running:

1) First, when the standard deviation of all the subclasses ($\sigma_i(C_j)$) is smaller than a predetermined parameter, i.e., the average standard deviation ($\bar{\sigma}_i$) for all the original classes, the clustering stops. This strategy assumes that only a few classes displaying significant spectral variations among he labeled cases, are to be generated. In other words, $\bar{\sigma}_i$ is the controlling parameter. Though the value of this predefined parameter can be manually set, setting it too

small will lead to too many subclasses (substrata) to be generated, which may make the classifier over-fitting or having too many noises and thus lacking prediction power. In the current case study, we compared a few results by setting different values for the parameter (i.e., $\bar{\sigma}_i$, $1.5\times\bar{\sigma}_i$, and $0.5\times\bar{\sigma}_i$) and found that $1.0\times\bar{\sigma}_i$ performed best in terms of the classification accuracy test. However, the trial-by-error method for setting up a value for $\bar{\sigma}_i$ is neither robotic nor the best for accuracy assessment. Future efforts should be made to explore better strategies for setting up the parameter.

2) Second, when the number of cases in a subclass is smaller than a predefined parameter, *MNC* (minimum number of cases), the analysis also stops. For the similar reason, a substratum containing too few cases in the training dataset will also lead to over-fitting or having too many noises.

To measure the similarity between any given case and a substratum from LCCs, a set of membership functions are defined based on each SS space of each SLCC. We first define a membership function for each variable for each substratum and then combine the effect of all the variables to make an overall determination function given by band weight vector ω_i. A trial-and-test method is employed to set the suitable weight vector. The membership function defined to measure the similarity of a given object to a substratum for a single variable takes the mean value and the standard deviation of the variable from the substratum labeled cases in the training dataset. The empirical parameter that affects the measured similarity value is the overlapping coefficient (β). As can be seen from Equation 6, the increase of β also increases the similarity measure value. While increasing β may lead to more cases to be classified, misclassification may occur when doing so. On the contrary, decreasing β may lead to some cases that actually belong to a substratum get lower similarity values and thus may be misclassified to other classes. In the case study, we adopt a value of 3 for β . This value considered the statistical properties (mean value and standard deviation) of the proximity between different classes and had the best accuracy when compared to other settings.

Although a moderate classification accuracy is obtained by the spectral substratum classifier in our case study, it is one of the best classification results among the present literature, considering the complicated vegetation cover and strong human influences in this region. In addition, the classifier is also easy to build and could be widely applicable to different environment conditions. Therefore, we suggest that the spectral substratum classifier should be further tested to extract information from remotely sensed images in other heterogeneous regions.

REFERENCES

Černá, L. and Chytrý, M., 2005, Supervised classification of plant communities with artificial neural networks. Journal of Vegetation Science, 16(4), pp. 407-414.

Cristianini, N. and Shawe-Taylor, J., 2000, An introduction to Support Vector Machines and other kernel-based learning methods. Cambridge University Press.

De Leeuw J., Liu, X., Schmidt, K., Skidmore, A.K., Jia, H. And Yang, L., 2006, Comparing accuracy assessments to infer superiority of image classification methods. International Journal of Remote Sensing, 27(1), pp. 223-232.

Du, C.J., Sun, D.W., 2008. Multi-classification of pizza using computer vision and support vector machine. Journal of Food Engineering 86(2), pp. 234-242

Gustavo, C.V., and Lorenzo, B., 2009. Kernel Methods for Remote Sensing Data Analysis. Wiley, Singapore, pp.125-148.

Triepke, F.J., Brewer, C. K., Leavell, D. M., Novak, S. J., 2008. Mapping forest alliances and associations using fuzzy systems and nearest neighbor classifiers. Remote Sensing of Environment, 112(3), pp. 1037-1050.

Laba, M., Ogurcak, D., Hill, E., Fegraus, E., Fiore, J., Degloria, S.D., Gregory, S.K. And Braden, J., 2002, Conventional and fuzzy accuracy assessment of the New York Gap Analysis Project land cover map. Remote Sensing of Environment, 81(2-3), pp. 443-455.

Langley, S.K., Cheshire, H.M., Humes, K.S., 2001. A comparison of single date and multitemporal satellite image classifications in a semi-arid grassland. J Arid Environ., 49(2), pp. 401-11.

Lo, C.P. And Choi, J., 2004, A hybrid approach to urban land use/cover mapping using Landsat 7 Enhanced Thematic Mapper Plus (ETM +) images. International Journal of Remote Sensing, 25(14), pp. 2687-2700.

Nordberg, M.L., Evertson, J., 2003. Vegetation index differencing and linear regression for change detection in a Swedish mountain range using Landsat TM and ETM+ imagery. Land Degradation & Development, 16(2), pp. 139–149.

Sha, Z., Bai Y., Xie, Y., Yu, M., Zhang, L., 2008. Using a hybrid fuzzy classifier (HFC) to map typical grassland vegetation in Xilin River Basin, Inner Mongolia, China. International Journal of Remote Sensing 29(8), pp. 2317-2337.

Sohn, Y., Rebello, N.S., 2002, Supervised and unsupervised spectral angle classifiers. Photogrammetric Engineering and Remote Sensing, 68(12), pp. 1271–1280.

Xie, Y., Sha, Z., and Bai, Y., 2010. Classifying historical remotely sensed imagery using a tempo-spatial feature evolution (T-SFE) model. ISPRS Journal of Photogrammetry and Remote Sensing, 65(2): 182-190.

Xu, M., Watanachaturaporn, P, Varshney P.K., et al., 2005. Decision tree regression for soft classification of remote sensing data. Remote Sens Environ., 97(3), pp. 322–36.

IMAGE TEXTURE PRESERVATION IN SPECKLE NOISE SUPPRESSION

A. Shamsoddini [a,*], J.C. Trinder [a]

[a] School of Surveying and Spatial Information Systems, University of New South Wales, Sydney, Australia- a.shamsoddini@student.unsw.edu.au, j.trinder@unsw.edu.au

KEY WORDS: SAR, Image, Radar, Texture, Edge, Detection, Model

ABSTRACT:

Speckle noise which occurs due to the coherent imaging system is the best known problem of SAR images and in turn, affects classification, change detection, biomass estimation and interpretation results. Several adaptive filtering methods have been documented to deal with this issue, such as Kuan, Lee, MMSE and Frost filters. These filters do not consider the level of homogeneity in the intensity of the pixels. For this reason, they degrade the spatial resolution of image and smooth details, while significantly decreasing the speckle noise level. There are other filters such as Enhanced Lee and Gamma Map that utilize the level of homogeneity, but they cannot adequately suppress speckle noise. Moreover, pixels whose coefficients of variation are near to maximum and minimum threshold values are not correctly filtered using these filters. In addition to these weaknesses, pixels surrounding a point scatterer are also treated as point scatterers due to shortcoming of the method of evaluating the coefficient of variation for differentiating between them and the point scatterer. We have developed a new method based on the homogeneity level for speckle noise suppression and simultaneously edge and feature preservation. Also, an algorithm has been proposed based on local statistical information to filter the pixels surrounding point scatterers. The results show an improvement in speckle reduction and texture preservation as well as reduction in the number of unfiltered pixels.

1. INTRODUCTION

Speckle noise, also referred to as 'speckle' is common to all imaging systems which utilize a coherent mechanism to acquire images, and SAR images are no exception (Bamler, 2000). In coherent systems, backscatter signals add to each other coherently and random interference of electromagnetic signals causes the speckle noise to occur in the image (Saevarsson et al., 2004). In fact, speckle is multiplicative noise that alters the real intensity values of features in a scene (Dong et al., 2001). Hence, speckle reduces the potential of SAR images to be utilized as effective data in remote sensing applications such as classification and segmentation, change detection, biomass estimation and interpretation, due to a degradation in appearance, quality and the recorded power of returns (Ali et al., 2008; Lee and Pottier, 2009). For this reason, speckle reduction becomes one of the more important tasks in radar remote sensing.

The main requirements that speckle suppression methods must meet are speckle reduction, and edge or texture preservation (Dong et al., 2001). In homogeneous areas filtering should only reduce the speckle noise level. A minimum unbiased estimator such as mean filter or box filter can perform very well and efficiently reduce speckle noise level over these areas (Lopes et al., 1990 b). Conversely, in the more heterogeneous areas, an ideal filter should suppress speckle noise and simultaneously preserve the edges and features, so a mean filter is not reliable for this type of data. According to these considerations, a good adaptive filter should have two important characteristics; first it should use an efficient discriminator to separate the speckle from the textural information and secondly, the filter should adaptively deal with speckle noise based on the type of speckle noise model which it follows (Lopes et al., 1990 b).

In general, speckle noise filters are grouped into two main categories:

- Statistical filters that use a priori statistical knowledge about speckle noise, the most common being Lee (Lee, 1981), Frost (Frost et al., 1982), Kuan (Kuan et al., 1985). These filters smooth speckle adequately, but they do not preserve details efficiently. Other statistical filters maintain feature information at the cost of poor speckle noise reduction, such as the Gamma Map (Lopes et al., 1990 a) and Enhanced Lee (Lopes et al., 1990 b) filters while all of the mentioned filters are based on speckle models. In addition, the latter filters are not able to filter large parts of images where the coefficient of variation is weak as explained later. There are other statistical filters such as mean and median filters which are not based on speckle models.
- frequency domain methods, such as Wavelet and Fourier transformations (Dong et al., 2001; Saevarsson et al., 2004; Maycock et al., 2007). These filters are not based on speckle models.

In this paper we aimed to develop a filtering method that can reduce the speckle noise and at the same time preserve the edges and features to acceptable levels.

2. SAR FILTERING CONSIDERATIONS

According to Lopes et al. (1990 b) common adaptive statistical filters have been developed based on the multiplicative noise model that assumes backscatter from a pixel originates from a large number of scatterers with independent phase and amplitude. This is not the case for built-up areas. Moreover, for the edges and some textured areas where details are smaller than the spatial resolution, the multiplicative noise model is unsatisfactory. Hence, for these two situations these filters are not efficient. On the other hand the filters mentioned above are based on using the local coefficient of variation, which is the ratio of standard deviation to the mean of pixels. This is known

* Corresponding Author

to be an efficient index of the homogeneity level of pixels over an image, but not a good textural measure as are second order statistical indicators such as the variance (Paudyal et al., 1995). In addition to these considerations there are some other shortcomings of filters using coefficient of variation as follows:

1- Using the coefficient of variation for the pixels that surround a point backscatterer is not reliable, because the coefficients of variation of these pixels are large, since their coefficient of variation is affected by the central pixels which are expected to be the point scatterers. This shortcoming causes these pixels to be inadequately filtered compared with the point scatterers. Therefore, applying a robust algorithm to deal with this problem is a demanding task.

2- Pixels with coefficients of variation near to C_{max}, the maximum coefficient of variation, are not filtered. Also pixels with coefficients of variation near to C_u, the averaged coefficient of variation over homogeneous areas, are averaged; although they are not classified as homogeneous areas, the averaging process will cause details to be lost.

3- For most statistical filters, averaging the pixels in heterogeneous areas with edges can lead to errors in filtered pixel values, because pixels with different noise models are combined in the averaging process.

3. METHODOLOGY

According to the previous section, in order to reduce or remove the above problems the following tasks are required; (i) A more robust criterion to discriminate different parts of image must be developed, not only based on homogeneity, but also according to textural features, (ii) The averaging of pixel values should be based on pixels whose speckle noise models are similar, and (iii) The development of an algorithm that deals with pixels which surround point scatterers or have homogeneity levels near to the maximum coefficient of variation. The method developed in this paper has been based on the determination of four thresholds from a standard deviation map derived within a 5 ×5 window. The method can be extended to larger windows.

3.1 Textural Criterion

In this study we have used edge detection masks to generate a new criterion for separating different textural areas in a SAR image. Considering a 5 × 5 window, it is possible to divide this window into nine 3×3 sub-windows corresponding to nine geographic directions. The mean value for each sub-window, which is called a sub-mean, is calculated and four 3×3 edge detection filters are separately scanned over the sub-means. Then, the results are summed and set to absolute values. This process results in 4 numbers whose standard deviations can provide textural information for different parts of a SAR image. The standard deviation map can be used as a textural criterion. The edge detection filters used are as follows:

$$\begin{bmatrix}-1 & 0 & 1\\ -1 & 0 & 1\\ -1 & 0 & 1\end{bmatrix},\begin{bmatrix}0 & 1 & 1\\ -1 & 0 & 1\\ -1 & -1 & 0\end{bmatrix},\begin{bmatrix}1 & 1 & 1\\ 0 & 0 & 0\\ -1 & -1 & -1\end{bmatrix},\begin{bmatrix}1 & 1 & 0\\ 1 & 0 & -1\\ 0 & -1 & -1\end{bmatrix}$$

According to Lee and Pottier (2009), these filters are affected by speckle noise less than other filters such as the Sobel filter.

3.1.1 Areas without Edges

Since no significant edges or textural features exist in a homogeneous area, except for some isolated pixels with very high or low values, it is possible to select the average value of the standard deviation map, V_{NE}, as the threshold. The areas with the standard deviation map values below the threshold contain no significant feature. Isolated points and their surrounding pixels will have relatively high standard deviation values compared to other pixels. In order to reduce the number of these pixels that may be filtered during the filtering process, we defined a second threshold, $V_{NE\text{-}max}$, which is the maximum value of the standard deviation map over the homogeneous area. Since using the maximum value results in some edges to be smoothed over edge areas, in order to reduce this problem, it is possible to select an area that has no point scatterers, where the standard deviations follow irregular curves, over edge areas. The average standard deviation of this area is the second threshold for the homogeneous area with a value between V_{NE} and maximum value of standard deviations. In summary, the non-edged area is divided into two different sub classes using two thresholds.

3.1.2 Edge Areas

The second class includes pixels that include edges and textural information. The low threshold of this class is $V_{NE\text{-}max}$ which is the high threshold of the previous class. The high threshold of this class, $V_{E\text{-}max}$ can be the maximum value in the standard deviation map over the area that includes edges and textural information. However, in order to decrease speckle noise level more over the heterogeneous area, it is better to select the average value of the standard deviation map over point scatterer areas as the high threshold for this class. The map of standard deviations over these areas appears as circular shapes, or closed curves.

3.1.3 Isolated Point Targets

The third class covering the remainder of image represents the point scatterer pixels and their neighbours. These pixels appear as closed curves and circular shapes in the standard deviation map and have the highest values. Figure 1 shows a part of standard deviation map including the different classes.

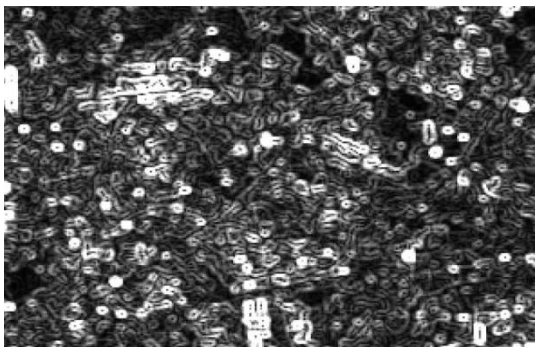


Figure 1. Standard deviation map for a part of study area; the brightest closed curves represent features, opened curves are edge areas and dark parts indicated homogeneous areas.

3.2 Filtering Scenarios

Since there are three different classes in terms of textural information over the images with different homogeneity levels, according to Lopes (1990 b) we need to use different scenarios for these different classes.

3.2.1 Non-edge Class Filtering

According to previous section this class should be divided in two sub classes. The first sub class includes pixels with standard deviation values less than or equal to $V_{NE,}$ which means that there is no textural information over this sub class. Hence, a minimum variance unbiased estimator can efficiently reduce the speckle noise level over this sub class without considering the textural information. The second sub class that comprises pixels with standard deviation values between V_{NE} and $V_{NE\text{-}max}$, describes isolated point scatterers. In this case there are two groups of pixels (i) pixels whose coefficients of variation are less than or equal to C_{max} , the coefficient variation over heterogeneous area, and (ii) pixels with coefficients of variation higher than C_{max}.

For the first group, the mean value of pixels within the selected window is used as the filtered pixel value because they are considered to be in the non-edge class and using the mean value does not degrade the spatial information. The second group represents isolated points and their surrounding pixels. According to section 2, one of the most important problems with the existing filters is that they consider the neighbouring pixels of point scatterers in the same way as the point scatterers themselves, since the coefficient of variation is not reliable for these pixels and hence is unable to separate them from the central point scatterer. In order to solve this problem and to filter these pixels we developed an algorithm that is called 'point scatterer discriminator'. This algorithm is based on the assumption that the difference of pixel values between point scatterers and their neighbours is high. After labeling a pixel as a candidate point scatterer, having a coefficient of variation higher or equal to C_{max}, a 3×3 window is centered on this point. Then, the maximum and minimum pixel values of this window are selected and the following equation is executed on all pixels within the window:

$$D = \frac{DN_{max} - DN_{ij}}{DN_{max} - DN_{min}} \qquad (1)$$

Where DN_{max}= the highest value within selected window

DN_{min}= the lowest value within selected window

DN_{ij}= the pixel value of pixel (i,j)

Then the median and mean values for the matrix of the differences, D, are calculated and the larger value, M, is used to make a decision about the central pixel. If the central value of the matrix of differences is less than M then this pixel is known as a point scatterer, otherwise the pixels whose difference values are more than or equal to M are selected and the coefficient of variation for selected pixels is calculated. As mentioned this sub class is not expected to include textural information except point scatterers. For this reason, the pixel is known as a point scatterer provided its coefficient of variation is higher than C_u, the averaged coefficient of variation over a homogeneous area, otherwise the mean value of the selected pixels is assigned as a filtered pixel value. Some isolated points with very low pixel values derived from this algorithm may also be preserved. These points are not recognized as point scatterers in the first step of this algorithm; however, the second and third steps can solve this problem.

3.2.2 Edge Class Filtering

Filtering the image over this class is more complicated than the other classes because these areas include textural information such as edges and built-up areas. For this reason, using the mean value over this class causes smoothing of the textural information and degrading of the image details; however, there will still be some homogeneous areas within this class that should be smoothed using a mean filter. Since the mean value of the standard deviation map over the point scatterer areas is used for $V_{E\text{-}max}$, there will be some point scatterers over this class that should be preserved. According to these considerations there are three types of pixels, those for which C is less than or equal to C_u, those for which C is between C_u and C_{max} and those whose coefficients of variation are more than C_{max}.

According to Lopes et al. (1990 b), pixels whose coefficients of variation are less than or equal to C_u follow a fully developed speckle noise model, and should be averaged. Since within the edge classes the pixel values vary in terms of homogeneity level, averaging all pixels should result in a loss of detail. So, it is necessary to select only pixels whose coefficients of variation are less than or equal to C_u for averaging. For the pixels whose coefficients of variation are higher than C_{max}, the point scatterer discriminator algorithm is used, thus preserving the point scatterers and their neighbours.

The most complicated filtering in this class is on the pixels whose coefficients of variation are between C_u and C_{max} because pixels in this category display edges and more textured areas. On one hand using simple averaging for these pixels is unreliable because of the high variability among the pixels. On the other hand, even if we utilize averaging using only pixels whose coefficients of variation are between C_u and C_{max}, it will introduce errors because they are not the result of fully developed speckle model. Hence, weighted averaging using more similar pixels in terms of homogeneity is more reliable for the filtering.

In order to deal with filtering of this part, after selecting pixels whose coefficients of variation are between C_u and C_{max}, the following equation called homogeneity likelihood (HL) is applied to find pixels of similar homogeneity within a window with respect to the central pixel:

$$HL = \left| \frac{C_c - C_{ij}}{C_{max} - C_u} \right| \qquad (2)$$

Where C_c= coefficient of variation for central pixel

C_{ij}= coefficient of variation for pixel (i,j) within the window

This index shows the similarity between neighbouring pixels with no fully developed speckle model and the central pixel. The lower the value of a pixel, the higher the similarity with the central pixel in terms of speckle model. Then this index is used to weight pixels using the following expression:

$$\begin{cases} w_{ij} = \exp(-HL) & \text{if } C_u < C_{ij} < C_{max} \\ w_{ij} = \exp(-\infty) & \text{if } C_{ij} <= C_u \text{ or } C_{max} <= C_{ij} \end{cases} \qquad (3)$$

And in order to normalize the weighting factors, we have:

$$W = \sum_i^m \sum_j^n w_{i,j} \qquad (4)$$

Then the weighted mean value is calculated as follows:

$$\bar{Z}_w = \frac{\sum_i^m \sum_j^n w.Z_{i,j}}{W} \qquad (5)$$

Where $Z_{i,j}$ is the pixel value (i,j)

After calculating the weighted mean, the coefficient of variation is calculated for the selected pixels. If this value is less than or equal to C_u, the filtered value is equal to weighted mean value, while if it is higher than C_{max} then the original value is preserved. For the pixels whose coefficients of variation are between C_u and C_{max}, the filtering method is as follows:

$$\text{Filtered pixel} = \bar{Z}_w \times B + Z \times (1-B) \qquad (6)$$

B is calculated according to the following equation:

$$B = \exp(-K \times N) \qquad (7)$$

Where K = damping factor

And N is calculated as follows:

$$N = \frac{C_{Sij} - C_u}{C_{max} - C_{Sij}} \qquad (8)$$

Where C_{Sij} = coefficient of variation for selected pixels

It is apparent that equations (6), (7) and (8) are similar to the equations that were proposed for enhanced Lee filter. It means that the more heterogeneous the pixels, the less filtering. However, there are some differences with the enhanced Lee filter including using the weighted mean instead of simple averaging, and applying the coefficient of variation for the selected pixels instead of calculating this value for all pixels within the window. There are some advantages in applying these changes. First, as mentioned in section 2, pixels whose coefficients of variation are close to C_u are averaged; however, their coefficients of variation are higher than C_u and their speckle model is not fully developed. Through using the weighted mean, this problem is removed because if B equals 1, then the filtered value is set to the weighted mean based on the noise model similarity. Moreover, for the pixels whose coefficients of variation are close to C_{max} the Enhanced Lee filter treats them as point scatterers, while the modified method is able to filter them through calculating the coefficient of variation for the selected pixels.

3.2.3 Point Scatterer Class Filtering

Pixels that have values more than or equal to $V_{E\text{-}max}$ in the standard deviation map are categorized as point scatterer candidates because some of them are pixels surrounding point scatterers. Therefore, it is necessary to use the point scatterer discriminator algorithm to find which pixels are point scatterers.

4. FILTER ASSESSMENT

There are several methods to assess the filtered image quantitatively according to different aspects such as noise reduction, edge preservation, feature preservation (Sheng and Xia, 1996). The results of these different measurements can be contradictory. Hence, different assessment methods should be used to find the optimum tradeoff among the different aspects of image quality assessment (Qui et al., 2004).

4.1 Equivalent Number of Looks (ENL)

This index is calculated using the following equation (Gagnon and Jouan, 1997):

$$ENL = \left(\frac{mean}{standard\ deviation}\right)^2 \qquad (9)$$

The higher ENL value for a filter, the higher efficiency in smoothing speckle noise over homogeneous areas.

4.2 Speckle Suppression Index (SSI)

This index is based on the equation as follows:

$$SSI = \frac{\sqrt{var\ (I_f)}}{mean\ (I_f)} \times \frac{mean\ (I_o)}{\sqrt{var\ (I_o)}} \qquad (10)$$

Where I_f = filtered image

I_o = noisy image

This index tends to be less than 1 if the filter performance is efficient in reducing the speckle noise (Sheng and Xia, 1996).

4.3 Speckle Suppression and Mean Preservation Index (SMPI)

ENL and SSI are not reliable when the filter overestimates the mean value. We developed an index called Speckle Suppression and Mean Preservation Index (SMPI). The equation of this index is as follow:

$$SMPI = Q \times \frac{\sqrt{var\ (I_f)}}{\sqrt{var\ (I_o)}} \qquad (11)$$

And Q is calculated as follows:

$$Q = R + \left|mean\ (I_o) - mean\ (I_f)\right| \qquad (12)$$

Where $$R = \frac{\text{Max (mean}(I_f)) - \text{Min (mean}(I_f))}{mean\ (I_o)} \qquad (13)$$

According to this index, lower values indicate better performance of the filter in terms of mean preservation and noise reduction.

4.4 Edge-Enhancing Index (EEI)

This value indicates how much a filter is able to preserve the edge areas and is defined as (Sheng and Xia, 1996):

$$EEI = \frac{\sum|DN_{1f} - DN_{2f}|}{\sum|DN_{1o} - DN_{2o}|} \qquad (14)$$

Where, DN_{1f} and DN_{2f} = filtered values of the pixels on either side of the edge

DN_{1o} and DN_{2o} = original values of the corresponding pixels

EEI values are usually less than 1 and higher values indicate better edge preservation capability.

4.5 Image Detail-Preservation Coefficient (IDPC)

The correlation coefficient between original image and filtered image over fine details such as point scatterers is defined as IDPC (Sheng and Xia, 1996).

5. RESULTS

In order to test the Proposed algorithm, we used ground-range HH and HV polarized L-band magnitude ALOS data that were

extracted from SLC data with dimensions 2031×1936 pixels. These images cover some homogeneous areas such as water bodies, forests, agricultural lands and urban areas and HH polarized image is shown in figure 2.

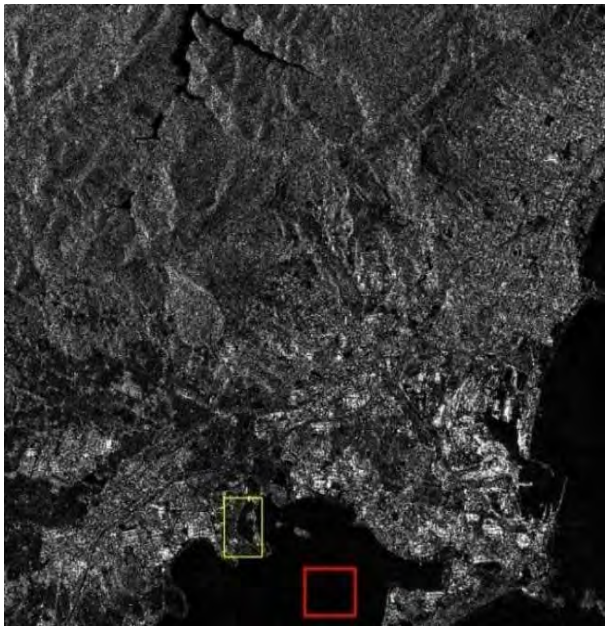


Figure 2. The HH polarized L band image; red rectangular shows the selected homogeneous area, yellow rectangular represents the selected edged area

5.1 Speckle Reduction

For the assessment of the performance of the filters to suppress speckle noise over selected homogeneous area, we used the three indices shown in table 3.

Filter	image	Mean ($\times10^{-3}$)	SD ($\times10^{-3}$)	ENL	SSI ($\times10^{-3}$)	SMPI ($\times10^{-3}$)
Noisy image	HH	101.30	53.3	----	----	----
	HV	28.70	14.8	----	----	----
Lee	HH	101.31	16.8	36.37	315	8.2
	HV	28.70	4.3	44.55	291	7.5
Kuan	HH	101.31	16.8	36.37	315	8.2
	HV	28.70	4.4	42.55	297	7.7
MMSE	HH	101.31	16.8	36.37	315	8.2
	HV	28.70	4.3	44.55	291	7.5
Frost K=1	HH	101.30	18.1	31.32	340	8.8
	HV	28.70	4.7	37.29	318	8.2
Enhanced Lee K=1	HH	100.95	28.5	12.55	537	16.8
	HV	28.60	8.2	12.16	556	16.2
Gamma Map	HH	98.70	20.9	22.30	402	20.2
	HV	27.96	5.76	23.56	399	20.1
Proposed K=1	HH	101.16	18.9	28.65	355	9.6
	HV	28.70	5.1	31.67	345	8.9

Table 3. Speckle noise reduction indices for the filtered images

As table 3 shows, the performance of Lee, Kuan and MMSE filters are very good for suppressing the speckle noise over the homogeneous areas whereas Enhanced Lee filter is not able to reduce the speckle noise efficiently. The Proposed method listed in the last line of table 3 shows comparable results in speckle noise reduction for HH polarized image.

5.2 Edge Preservation

In order to use EEI index, the edge between water body and land was selected. This area is shown within a yellow rectangle in figure 1. The results of this index for the filters are given in table 4. The best algorithm performance for the edge preservation with the highest EEI values is the Enhanced Lee and the Proposed method respectively. On the other hand, Frost, Kuan, Lee and MMSE filters are not able to preserve the edges. It is estimated that in filtered images derived using the Enhanced Lee filter and the Proposed method, edges are up to 2 times sharper than Lee, Kuan, MMSE and frost filters.

Filter	image	EEI ($\times10^{-3}$)
Lee	HH	396.7
	HV	262.8
Kuan	HH	448.1
	HV	289.5
MMSE	HH	361.9
	HV	244.5
Frost K=1	HH	313.4
	HV	254.8
Enhanced Lee K=1	HH	999.7
	HV	918.1
Gamma Map	HH	968.2
	HV	731.6
Proposed K=1	HH	999.0
	HV	833.2

Table 4. Edge index values for different filters

5.3 Preservation of Details

More than one thousands pixels representing significant features were selected separately over the two images and, the correlation between filtered and original images over the selected pixels was calculated. Table 5 presents the results of this index. The best feature preservation performance belongs to Proposed method and Enhanced Lee filter for which their index values show no variation for all features.

Filter	image	IDPC
Lee	HH	0.94
	HV	0.96
Kuan	HH	0.95
	HV	0.96
MMSE	HH	0.94
	HV	0.94
Frost K=1	HH	0.91
	HV	0.89
Enhanced Lee K=1	HH	1.00
	HV	0.99
Gamma Map	HH	0.98
	HV	0.98
Proposed K=1	HH	1.00
	HV	1.00

Table 5. IDCP of the filters over selected features

5.4 Filtering of Pixels

According to section 2, some pixels should be preserved and do not need to be filtered. On the other hand, a filtering method should filter all pixels where necessary. As mentioned earlier, some pixels surrounding features and point scatterers that should be filtered but are not filtered by the Enhanced Lee and Gamma filters because of the deficiency of the coefficient of variation in their location. We developed an algorithm to deal with this problem. In order to assess this algorithm, 100 pixels representing point scatterers were selected over these images and the coefficient of variation map assessed. Then the filtered pixels were divided by the corresponding pixels of original images within a 5×5 window. In this way, pixels whose values are 1 are categorized as unfiltered pixels. The results were given in table 6.

Filter	Image	Number of unfiltered pixels	Filtering performance (%)
Enhanced Lee K=1	HH	2229	11
	HV	2347	6.1
Gamma Map	HH	2162	13.5
	HV	2322	7.1
Proposed K=1	HH	1374	45
	HV	1459	41.6

Table 6. Number of unfiltered pixels over point scatterers

Table 6 reveals that the point scatterer discriminator algorithm can perform very effectively in compensating for the deficiency of calculating the coefficient of variation for the pixels which are near the point scatterers. As this table shows, filtering performance for the Enhanced Lee and Gamma filters over selected point scatterer areas are very poor as they are able to filter less than 14 percent of these pixels whereas the proposed filter increases the number of filtered pixel to 45 percent.

6. CONCLUSION

In this study a new algorithm based on coefficient of variation similarity and using a new criterion to segment different parts of the SAR image has been proposed. This method was compared to six common filters using different quantitative assessment methods. According to the assessments that were used in this study, some filters such as Lee, Kuan and MMSE filter perform very efficiently in dealing with the problem of speckle noise at the expense of smoothing features and edges. Some other filters such as Enhanced Lee and Gamma Map can preserve details very efficient, but they are not able to reduce speckle noise. In addition to this, the inadequacy of the coefficient of variation causes these filters to be unable to deal with the problem of speckle noise of the pixels surrounding point scatterers and fine features. Meanwhile, pixels whose coefficients of variation are close to C_u are averaged, while if they are higher than C_u they should be treated as pixels whose speckle noise model is not fully developed.

In this study we proposed a novel model to deal with these problems. As the results show, the proposed filtering method can perform acceptably well in speckle reduction and simultaneously edge and feature preservation. In addition to this, the point scatterer discriminator algorithm that was developed in this study and used in the structure of the proposed method can compensate for the deficiency in the coefficient of variation in separating between point scatterers or features and the pixels surrounding them. Finally, the proposed method is being examined to prove its validity for other types of data.

REFERENCE

Ali, S. M., Javed, M. Y., Khattak, N. S., Mohsin, A. & Farooq, U., 2008. Despeckling of Synthetic Aperture Radar Images Using Inner Product Spaces in Undecimated Wavelet Domain. *World Academy of Science, Engineering and Technology,* 38**,** 167-172.

Bamler, R., 2000. Principle of Synthetic Aperture Radar. *Surveys in Geophysics,* 21**,** 147-157.

Dong, Y., Milne, A. K. & Forster, B. C., 2001. Toward edge sharpening: a SAR speckle filtering algorithm. *Geoscience and Remote Sensing, IEEE Transactions on,* 39**,** 851-863.

Frost, V. S., Stiles, J. A., Shanmugan, K. S. & Holtzman, J. C., 1982. A Model for Radar Images and Its Application to Adaptive Digital Filtering of Multiplicative Noise. *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* PAMI-4**,** 157-166.

Gagnon, L. & Jouan, A., 1997. Speckle filtering of SAR images : A comparative study between complex-wavelet-based and standard filters. *SPIE proceedings series* 3169**,** 80-91.

Kuan, D. T., Sawchuk, A. A., Strand, T. C. & Chavel, P., 1985. Adaptive Noise Smoothing Filter for Images with Signal-Dependent Noise. *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* PAMI-7**,** 165-177.

Lee, J.-S., 1981. Speckle analysis and smoothing of synthetic aperture radar images. *Computer Graphics and Image Processing* 17**,** 24-32.

Lee, J.-S. & Pottier, E., 2009. *Polarimetric Radar Imaging from basics to applications* CRC Press, pp. 143-152.

Lopes, A., Nezry, E., Touzi, R. & Laur, H., 1990a. Maximum A Posteriori Speckle Filtering And First Order Texture Models In Sar Images. *Geoscience and Remote Sensing Symposium, 1990. IGARSS '90. 'Remote Sensing Science for the Nineties'., 10th Annual International.*

Lopes, A., Touzi, R. & Nezry, E., 1990b. Adaptive speckle filters and scene heterogeneity. *Geoscience and Remote Sensing, IEEE Transactions on,* 28**,** 992-1000.

Maycock, J., Hennelly, B. M., McDonald, J. B., Frauel, Y., Castro, A., Javidi, B. & Naughton, T. J., 2007. Speckle reduction using the discrete Fourier filtering technique. *Machine Vision and Image Processing Conference, 2007. IMVIP 2007. International.*

Paudyal, D. R., Eiumnoh, A. & Aschbacher, J., 1995. Textural information in SAR images for land-cover applications. *Geoscience and Remote Sensing Symposium, 1995. IGARSS '95. 'Quantitative Remote Sensing for Science and Applications', International.*

Qiu, F., Berglund, J., Jensen, J., Thakkar, P. & Ren, D., 2004. Speckle Noise Reduction in SAR Imagery Using a Local Adaptive Median Filter. *GIScience & Remote Sensing,* 41**,** 244-266.

Saevarsson, B. B., Sveinsson, J. R. & Benediktsson, J. A., 2004. Combined wavelet and curvelet denoising of SAR images. *Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. Proceedings. 2004 IEEE International.*

Sheng, Y. & Xia, Z., 1996. A comprehensive evaluation of filters for radar speckle suppression. *Geoscience and Remote Sensing Symposium, 1996. IGARSS '96. 'Remote Sensing for a Sustainable Future.', International.*

TEXTURE ANALYSIS TO IMPROVE SUPERVISED CLASSIFICATION IN IKONOS IMAGERY

A. N. Tassetti [a,*], E.S. Malinverni [a], M. Hahn [b]

[a] DARDUS, Polytechnical University of Marche, Ancona, Italy - (n.tassetti, e.s.malinverni)@univpm.it
[b] Dept. of Geomatics, Computer Sciences and Mathematics, Stuttgart University of Applied Sciences, Stuttgart, Germany- michael.hahn@hft-stuttgart.de

KEY WORDS: IKONOS, Land Cover, Training, Classification, Texture, Feature

ABSTRACT:

The most extensive use of Remote Sensing data is in land cover/land use (LCLU) studies by means of automated image classification. The general objective of this research is to develop an automatic pixel-based classification methodology with the aim to produce a Regional land use map congruent with the CORINE Land Cover legend. Starting point are detailed ground data, already gathered fostering interoperability among several Regional bodies' DBs and high resolution multi-spectral IKONOS imagery.
In the light of land mapping, there are two main features related to IKONOS imagery: lack of spectral information (4 spectral bands) and high spectral variability (high spatial resolution). This results in problems in terms of class information extraction especially using pixel-based image classification methods in which spatial information existing between a pixel and its neighbours is not used. To overcome these deficits, the use of vegetation indexes (NDVI feature and TDVI masks) and texture (GLCM and edge-density features) is investigated with respect to its impact on land cover/land use classification.
The developed spectral/textural classification schema is compared with the classical approach using only spectral information. An accuracy assessment is carried out which shows that image data with 4 IKONOS spectral bands plus NDVI band plus 6 texture bands achieve an accuracy of 80.01% compared to 63.44% of accuracy achieved by using the few spectral bands only. Furthermore it allows the discrimination of 10 CLC classes.
Experimental results show how, starting from available but also binding data (IKONOS imagery and available Regional ground data), a classification schema can be developed with enhanced performance and strong relation to the specific setup.

1. INTRODUCTION

This work is part of a wider project whose general objective is an automatic pixel-based classification methodology aimed at producing a regional CORINE Land Cover (CLC) land use map. Starting points are high resolution multi-spectral IKONOS imagery and ground data already gathered in previous works (Marcheggiani et al., 2008). The images, provided by Marche Region Institution, are mono-temporal (June 2006) and with only 4 spectral bands. The ground data, owned by different bodies of Marche Region public administration at regional level, are large and detailed. Consequently the main goal of this work is the development of a classification schema that can take advantage of all the available data in Region's possession (IKONOS images and Ground data) to investigate the possibility of a land cover mapping trustable enough to be a permanent monitoring service of Marche region territory, congruent with the land use oriented European trend (CLC legend).
In the light of land mapping, there are two main features related to IKONOS imagery: lack of spectral information and high degree of spectral variation due to the high image spatial resolution. This results in problems in terms of class information extraction especially using pixel-based image classification methods in which spatial information existing between a pixel and its neighbours is not used. To overcome these deficits and achieve reliable and accurate results, spectral and texture information (GLCM and edge-density features) are combined together in the proposed classification schema.

A fundamental goal of this research is in fact to explore the image texture information and how to combine it with the spectral signatures to do image analysis. Moreover the use of vegetation indexes (NDVI feature and TDVI masks) is investigated with respect to its impact on land cover/land use (LCLU) classification.
The study case focuses on the north-eastern part of the Marche region, belonging to the Ancona Province. It covers an area of approximately 80 km^2, comprising urban and rural landscape and natural Mediterranean environment, among which the Conero Mountain Natural Park have to be mentioned. Figure 1 gives an overview of the study image and its geographic location.

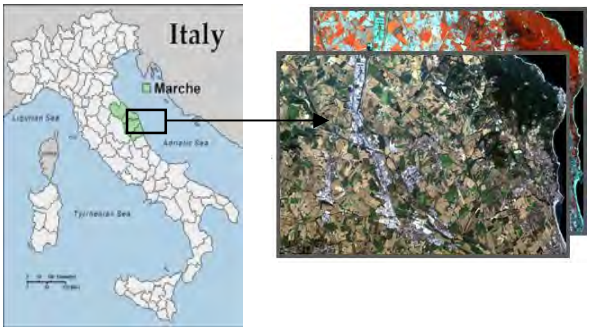


Figure 1. Map of Italy and Marche Region (left), test image in RGB and False Color composition (right).

* Corresponding author

2. METHODOLOGY

The proposed classification schema is pointed out below in Figure 2.

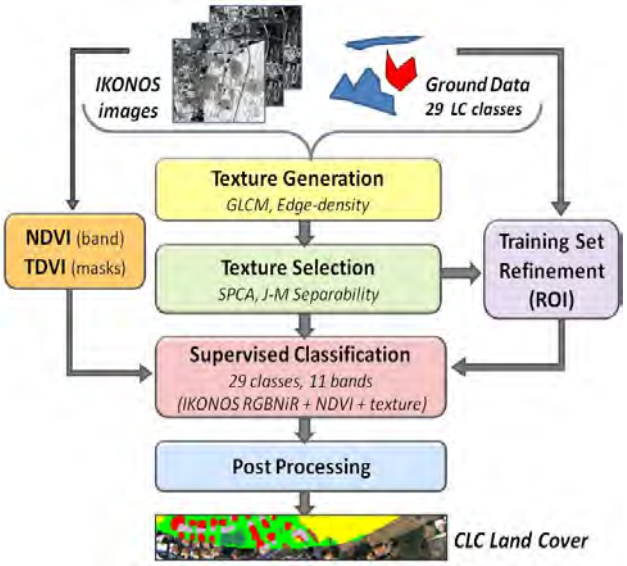


Figure 2. Workflow diagram describing the developed classification schema

The input data are multi-spectral IKONOS images and available ground samples gathered from different Information Systems owned by Marche Region and organized in independent learn and control samples. They are organized in 29 LC classes according to their cover type and they mostly differ from the CLC land use oriented nomenclatures. This means that, to match the CLC legend, they must be grouped together according to their use (agriculture, settlement etc.) that often doesn't fit their spectral response, making the supervised training process difficult. In this context another objective of this research is to investigate the best way to combine the 29 LC classes into CLC nomenclatures.

Vegetation indexes are investigated with different purpose: NDVI to generate an additional band and TDVI to build binary masks trough which three supervised classifications are carried out independently.

Texture features can be generated after having chosen the spectral bands to process and set the window size parameter (semivariogram/correlogram-guided texture feature generation). In particular 28 different texture features are generated according to two different texture approaches (GLCM and edge-density features). By means of Standardized PCA these 28 texture features are reduced or selected according to their loading factor, augmented by the four spectral bands and the NDVI feature and grouped into three different feature sets that differ only in the texture subset selection (IKONOS RGBNiR + NDVI + texture). According to the Jeffries-Matusita (J-M) average separability distance the best suited feature set can be chosen. It is used to refine the training data and to run the three TDVI masked supervised classification. A post processing step is needed to match the CLC legend and improve the spatial consistency of the pixel-based classification.

Hereafter the different steps shown in Figure 2 are explained in more details.

2.1 Semivariogram/correlogram-guided texture feature generation

Texture features could be theoretically calculated for each spectral band and for many window sizes but with the disadvantage of increasing the feature space dimensionality and redundancy. Some choices must be made.

2.1.1 Spectral band analysis

- AIM: to minimize feature space dimensionality by selecting optimal channels for texture measures.
- RESULT: Red and NiR bands are selected. They show highest variances for the different land covers and low correlation.

The following strategy to select the best band combination is chosen. The 29 LC classes samples are downgraded in 5 main cover classes and used as masks to compute covariance matrices. According to the different 5 LC main classes, the Red and the NiR band always correspond to the highest variances which indicate strongest texture features. Moreover they show low correlation.

2.1.2 Geostatistical correlogram/semivariogram analysis

- AIM: to investigate the optimum window size to use to generate texture features
- RESULT: 3x3,5x5 and 7x7 window sizes for the Red band, 3x3 and 5x5 for the NiR band

Semivariograms and Moran's *I* correlograms of all the 29 LC land cover classes are sequentially computed for lag distances increasing to 20 pixels. The radiometric spatial autocorrelation of the each particular LC class can be quantified in terms of the lag (*range*) that results in the maximum variability (*sill)* for the semivariogram and at the same time in a very close to zero Moran's *I* value. An example is displayed in Figure 3 and is related to one of the 29 LC classes (class 40305, Sparsely vegetated areas).

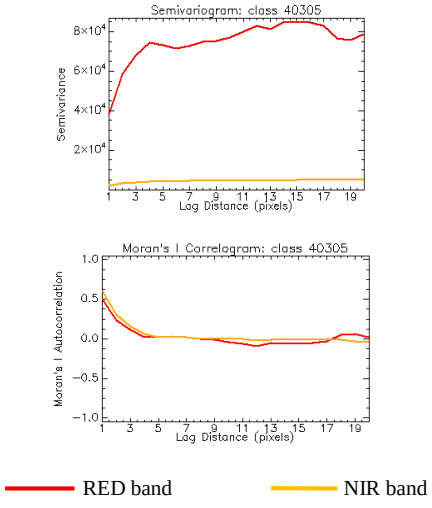


Figure 3. Red and NIR Semivariogram/Correlogram plots

Having a look to all the 29 pairs of LC class variograms and correlograms, they indicate that semivariances for Red band mostly start to saturate at a lag of 5 while some classes require 3 and 7 pixels of kernel size. Instead in the NiR band each land cover class reveal spatial correlation for lag distance of less than 5 pixels. Consequently 3x3, 5x5 and 7x7 are used as window sizes for the GLCM computation from the Red band, and two window sizes (3x3 and 5x5) are used to create NiR co-occurence features. The same window sizes are taken into account for the edge density image generation.

2.1.3 Texture feature generation

- AIM: generation of optimal texture features to add to the originals multispectral bands, before running the classification.
- RESULT: 20 GLCM features + 8 edge-density features = 28 texture features generated.

Two different kinds of texture are generated: Grey Level Co-occurrence (GLCM) and edge-density features.

***GLCM texture generation*:** As suggested by Hall-Beyer (2000) a combination of only four GLCM measures (Haralick, 1973) are selected to avoid that texture features are correlated with each other. In particular mean (MEAN), Variance (VAR), Entropy (ENT) and Homogeneity (HOM) are computed for the Red and NiR bands and five window size (respectively 3x3, 5x5 and 7x7 for the Red band and 3x3 and 5x5 for the NiR band) leading up to the generation of 4 x 5= 20 GLCM features. For reducing the degrees of freedom of the GLCM texture generation, the distance between pixels for the co-occurrence matrix computations is maintained constant at one and the average of the four main inter-pixel angles (0°, 45°, 90° and 135°) is used, based on the assumption that no land cover exhibits a preferential directionality. The gray scale quantization levels is set to 64: it allows to have a better computational and statistical performance and reduce processing time limiting the size of the co-occurence matrix to 64 x 64 (instead of 2048 x 2048 because of the radiometric resolution of 11 bit).

Edge-density texture generation: As done before, edge density maps are produced processing the Red and NiR bands. Firstly, each band is filtered using a Laplacian high pass filter; secondly, edges are found by thresholding the filtered image based on histogram interpretation. Finally, an average filter is used to produce the edge density map, counting the edge points in each position of the moving kernel and dividing the number of edge points by the window size. Two different thresholds for the Laplacian filter are selected interactively for each band (T_{R1} and$T_{R2,}$ T_{NiR1} and T_{NiR2}) and 9x9 and 15x15 are used as average filter's window sizes. 4 x 2 = 8 edge-density images are generated from the Red and NiR bands.

2.2 Texture selection

Before adding the collected 28 texture features as additional bands to the IKONOS imagery, they are investigated and sorted according to the standardized Principal Component Analysis (SPCA) with the aim to extract optimum linear combinations (Principal Components) of the original texture features that contain as much as possible variability of the original data.

2.2.1 Standardized Principal Component Analysis (SPCA)

- AIM: to reduce the generated texture features to the maximum number of uncorrelated data
- RESULT: 3 different feature sets are constructed:
  - *1st Feature set* : 4 GLCM PCs + 4 spectral bands + 1 NDVI
  - *2nd Feature set*: 6 GLCM&edge PCs + 4 spectral bands + 1 NDVI
  - *3rd Feature set*: 12 high "loading" features + 4 spectral bands + 1 NDVI

Assuming that the original texture features are more or less equally important, the problem to tackle is that the 28 texture variables have very different means and/or standard deviations. In this case a normalization is needed to avoid the importance of a variable being determined its variance that could dominate the whole covariance matrix and hence all the eigenvalues and eigenvectors. This standardization is done by running the Standardized PCA (SPCA) that equalizes dissimilar variations in the data set by using a correlation matrix instead of a covariance matrix (PCA). In particular the SPCA is performed two times (to the 20 GLCM features first and then to the 28 GLCM&edge features) to construct three different feature sets that later on must be investigated and selected before going on with the maximum likelihood classification process. Compromising three different selection guidelines (cumulative percentage, Scree plot and Kaiser's rule), the first four GLCM PC bands and six GLCM&edge PC bands are selected along with the original four spectral bands and the NDVI band to build the 1st and the 2nd feature set (to the amount of 9 and 11bands).Then, based on computed component loadings, the band combination (12 GLCM&edge bands) that have higher variance explained on various PCs is selected and added to the other 5 features (spectral bands and NDVI) to build the 3rd feature set (to the amount of 17 bands).

In order to guarantee that each source (spectrum, texture and NDVI) makes the same contribution to the feature space and avoid scale effects in the Maximum Likelihood statistic computation, each source of data is stretched from 0 to 1before running the classification schema.

2.2.2 Separability analysis

- AIM: to select the best suited feature set assessing class separability and expected classification errors for different feature combinations.
- RESULT: the second feature set (GLCM&edge) is selected because it corresponds to higher average separability of LC classes.

Average separability measures between each LC class are calculated using the three different built feature sets that include the 4 original image channels, the NDVI and different texture subsets (GLCM, GLCM&edge-density and "Loading"features). So, first, the pairwise J-M distances between each pair of classes is determined for all combinations of two, then the average J-M distance is computed for each class.
Figure 4 represents the average separabilities of the 29 LC classes as function of the average J-M distance.

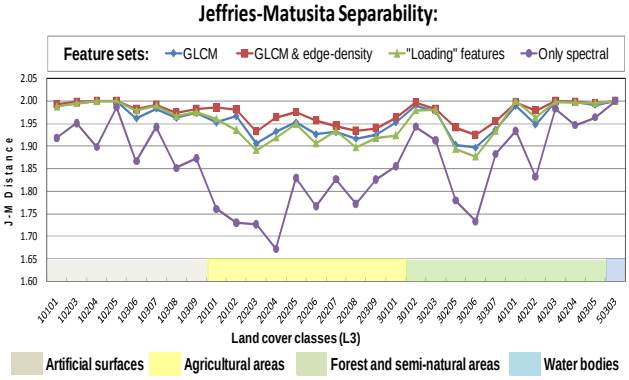


Figure 4. J-M Average separability for L3 cover classes

As shown in Figure 4, the second feature set (PCs from GLCM & edge-density features) is the best feature combination in order to separate the given LC classes because it shows mostly higher

values for the average J-M distances. The separability of clusters generated using multispectral bands in combination with the selected texture images, has improved especially for agricultural areas and some forest and semi-natural classes. Without texture the average J-M distance is often below 1.9 that is the threshold below which the separability is indicated to be poor.

2.3 Training set refinement (ROI)

- AIM: to improve ROI's representativeness.
- RESULT: class signatures are "cleaned" using the selected 2nd feature set and redefined as final ROIs.

Outlying pixels (in feature space) are deleted before computing the final class signatures. This can be done by self-classifying the training pixels according to the 2nd feature set. Misclassified pixels are excluded from the training set recalculating the final ROI (Region Of Interest) signatures only according to the well classified training pixels. In the Figure 5 is shown an example (class 40202) to summarize the "cleaning" workflow.

Original sample | Self-classified sample | "Cleaned" sample (ROI)

Figure 5. Class 40202 cleaning and ROI generation

Which kind of texture should be taken into account to improve the accuracyThe improvement in the training data can also be checked again in terms of statistical J-M separability: before the "cleaning" a lot of critical J-M distances (lower than 1.9) indicate classes non well separated with some overlaps in their density functions. Improvements are instead shown after the refinement: the number of critical class pairs decreases from 40 to 22 (highlighted in pink in Figure 6). Looking more closely at this J-M matrix (Figure 6), it is possible to investigate which LC classes are still not sufficiently separated in the given feature space and how to manage them. When the non separable LC class pairs belong to the same CLC class (level 1 or 2), it is not really a problem: according to the final CLC nomenclature, they be merged after the classification into more generalized CLC classes. Problems remain when it is not possible because the "critical" classes differ in the CLC level 1 itself. It is for example the case of the class 40305 (sparsely vegetated areas) not separable from the class 10308 (mix coverage buildings) even if belonging to different CLC level 1 class: Forest and semi-natural areas the former, Artificial surfaces the latter.
An expedient to overcome this problem is the use of vegetation indices.

2.4 Vegetation Indices

- AIM: to improve the classification accuracy and minimize the error matrix off-diagonal elements.
- RESULT: two vegetation indexes are generated:
  - NDVI to use as additional band in the selected feature set
  - TDVI to build three masks trough which run three supervised classifications.

TDVI (Bannari et al. 2002) is employed to develop thresholds useful to build binary masks (Figure 8) trough which three supervised classification processes are carried out independently.
Studying the TDVI histograms associated with the more critical classes, two thresholds (T_1=-1.2 and T_2=-0.6) are developed with the aim to identify pixels likely to belong to particular classes. For example in Figure 8 it is shown how T_2 can help to distinguish the mix coverage building class with value mostly below T_2 from the sparsely vegetated areas class with value mostly above T_2.

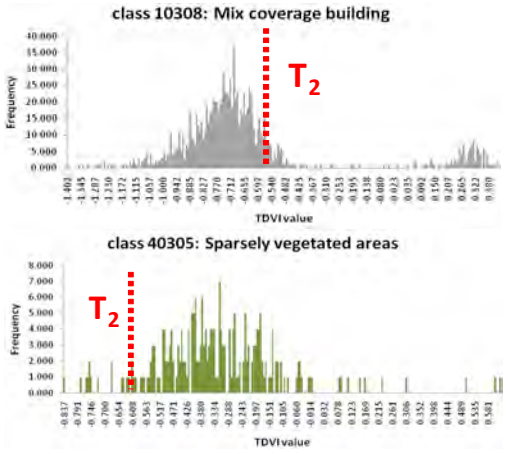


Figure 7. TDVI Histogram thresholding

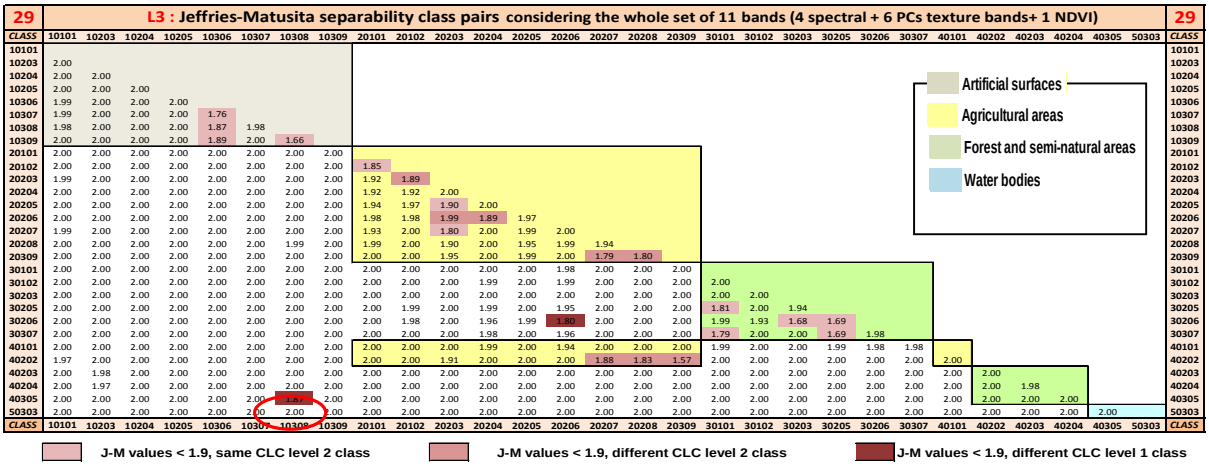


Figure 6. J-M separability class pairs distances after ROI "cleaning"

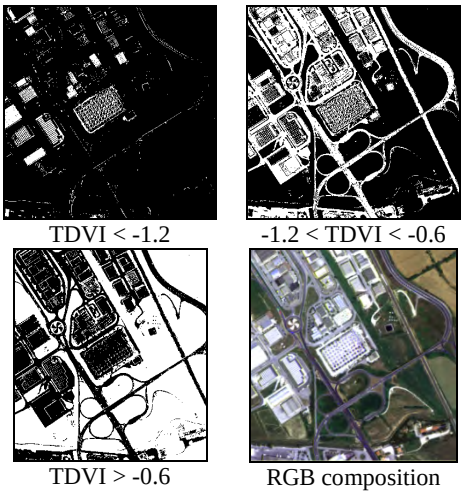


Figure 8. TDVI masks and RGB composition

After having defined the two thresholds, it is possible to study all the 29 TDVI class histograms to decide which sample use as ROI in each specific classification process. With these three ROI subsets three complementary Maximum Likelihood (ML) classification can be carried out independently.

2.5 Supervised classification

- AIM: to classify each pixel into one of the 29 LC classes.
- RESULT: merging of three complementary classification maps.

The three ML classifications are initially performed by default assuming that all the cover classes are equally likely. Then they are adjusted using a-priori information gathered by means of a ROC (Receiver operating characteristic) analysis linked to the LC ground data. By means of ROC curves is possible to visualize the performance of the classification method, in order to select proper decision thresholds providing the best classification with the minimum error rate. This ROC analysis is performed for each classification (3 times) and for all the ROI involved with the wise to take into account only the ground information allowed by the specific mask used. This a-prior information can give a crucial effect to classification results. In this way the ground data give again a powerful hint to drive the classification schema development.
After having performed the three ML classifications, the three outputs are merged into a single classified map by means of raster calculations.

2.6 Post- classification data manipulation

- AIM: to match the CLC legend and improve the spatial consistency of the pixel-based classification.
- RESULT: final CLC land cover map

The merged classified image is post-processed combining the 29 LC classes into10 generalized classes, according to the CORINE Land Cover nomenclature.
Finally post-classification techniques (majority analysis, sieving and clumpling) are applied to the CLC classified output in order to eliminate the 'salt and pepper' noise, removing gaps within areas covered by a predominant class.

3. RESULTS

In Figure 9 the performance of the developed spectral/textural classification schema is assessed drawing a comparison between the results obtained using only the spectral band and the improved results achieved integrating texture features (without and with the use of the TDVI mask).

		SPECTRAL	SPECTRAL + TEXTURE	
			without TDVI masks	with TDVI masks
Corine Land Cover Legend		Prod. Acc.	Prod. Acc.	Prod. Acc.
Code	DESCRIPTION	(Percent)	(Percent)	(Percent)
1.1	Urban Fabric	98.61	97.19	96.02
1.2	Industrial, commercial and transport units	39.45	60.21	73.95
2.1.1	Non-irrigated arable land	47.33	81.32	77.3
2.2	Permanent crops	30.29	64.25	68.25
2.3	Pastures	94.29	74.77	74.71
3.1	Forest	93.06	97.11	96.02
3.2	Scrub and/or herbaceous vegetation assoociations	62.06	91.21	88.19
3.3	Open spaces with little or no vegetation (beaches,dunes,bare rocks)	39.22	48.64	64.99
3.3.3	Sparsely vegetated areas	91.95	94.85	88.63
5.2	Marine waters	100	97.17	99.01
Code	Overall Accuracy (Percent)	63.44	74.38	80.07

Figure 9. Accuracy comparison (spectral versus. spectral+texture features)

Comparing the spectral and spectral/texture classifications in Figure 9, it is clear that spectral classification is better suited for those land use classes with a specific spectral response and well differentiated from the rest of the units, such as pastures (class 2.3) and marine waters (class 5.2). The distribution of grey levels in these two classes is very homogeneous, so they are more difficult to discriminate by texture methods. Rather, adding texture, their accuracy get worse. On the other hand, texture techniques are very efficient in classifying landscape units that contain a high spectral heterogeneity, such as permanent crops, scrub and/or herbaceous vegetation associations and non-irrigated arable land (class 2.1). These classes are not very accurate when classified using the spectral band only. For example, taking texture measures into account (without TDVI masks) the accuracy of the permanent crop is really improved (from 30.29% to 64.25%).
Similar explanation can be given for the industrial, commercial and transport units: the high edge density encountered in industrial area (class 1.2) allow to separate them from the other classes and especially from the open spaces with little or no vegetation (class 3.3) that were shown to be very spectrally similar. Regarding these two class an additional improvement is given by the mask use: class 1.2 can even reach an accuracy of 73.95%. while class 3.3 , although improving, still does not reach a satisfactory accuracy. This verifies the use of TDVI masks in the classification schema. However for this last class (class 3.3) problems can be caused above all by a bad training. In fact, having a close look at the particular samples generating the signature, for this class it is clear that they are mostly made up of beaches (very few samples for the other LC classes grouped into this class 3.3) and especially spectrally "mixed" beaches, as they can be in June in Italy because of beach umbrella and so on.
Another interesting aspect is that the integration of spectral and texture bands for classification has a synergic effect on the results, in some cases even improving the accuracy of both groups of classes (homogeneous and not). However, it is important to note that , according to the truth data available, the reported results refer to the inner truth areas of the texture units

and not to the borders between textures. Further work should be done to reduce the border effect.
A visual inspection of the final CLC map (Figure 10) confirms that the results of this developed classification schema are reasonably good. However an object-based post classification method is advisable.

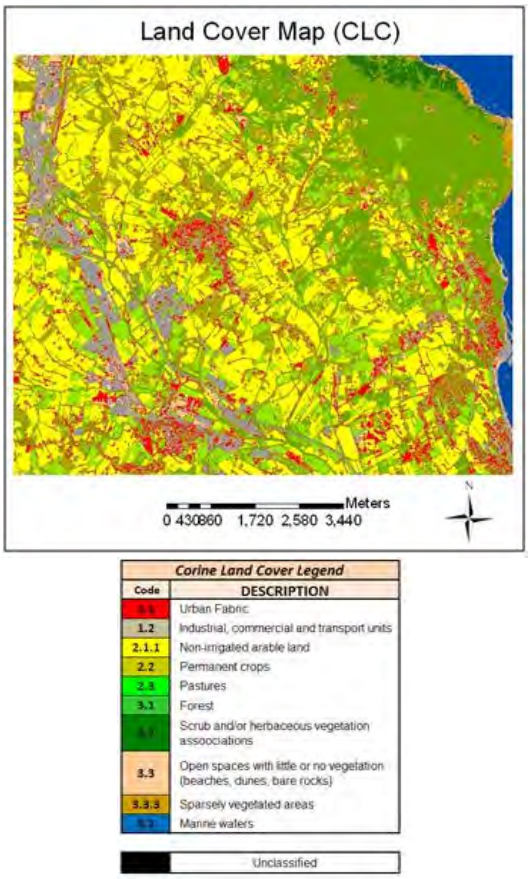


Figure 10. Final Corine Land Cover map with legend

4. CONCLUSION

The research shows how is possible to recycle and get benefit from large and detailed available ground information taking advantage from the IKONOS imagery and the potentiality offered today by remote sensing techniques.
To get satisfactory results it is necessary to develop a supervised classification schema integrating texture features.
The study confirms the utility of textural analysis to enhance the per-pixel classification accuracy. In particular it shows how is possible to extract texture features using second order GLCM statistics and edge-density images and how, after a targeted feature selection by means of the PCA, is possible to use them as additional bands in the classification schema. These new textures turn out to be useful auxiliary data especially for high resolution data sets suffering from high spectral heterogeneity. By incorporating these texture features in the classification schema, it is possible to achieve a higher classification accuracy compared to the classification of the original IKONOS image.
Particular improvements are shown especially in discriminating between agricultural species and semi-natural areas (e.g. open space with little or no vegetation). In particular permanent crops are impossible to discriminate without texture from the other classes
The overall accuracy increases to 80.0 % with a Kappa Coefficient of 0.7337 and the Producer's accuracies for the different classes increase as well.
Important to underline is that the use of texture features makes it possible to well-identify more CLC classes (10 CLC classes in the study case): thus it is fair to think that, increasing the amount of information extracted from the image, it is possible to reduce the support given by the photointerpreter in CLC map generation.
This work demonstrates the need of a spectral/textural image analysis for a more accurate land cover type discrimination when thematic classes are very heterogeneous (high within-class spectral variance) and spectral information is no longer a sufficient indicator for the classification.
Anyway no general rules can be recommended by this study for the texture measure selection: the most appropriate combination of texture features depends strongly on the surface properties of the land cover types of interest. What is found by this research is to optimize the window size according to the available ground data and also to choose the best feature set in terms of separability analysis (linked again to the collected ground data).

REFERENCES

Bannari, A., Asalhi, H., Teillet, P.M., 2002. Transformed Difference Vegetation Index (TDVI) for vegetation cover mapping. *International Geoscience and Remote Sensing Symposium*, Toronto, Ontario, Canada, paper I2A35-1508.

Hall-Beyer, M., 2000. GLCM texture: a tutorial. http://www.fp.ucalgary.ca/mhallbey/tutorial.htm. (accessed Nov.2008).

Haralick, R.M., Shanmugam, K., Dienstein, I., 1973. Textural features for image classification. *IEE Transactions on Systems, Man and Cybernetics*, vol.3, no.6, pp.610-617.

Marcheggiani, E.,Galli, A., Bernardini A., Malinverni, E.S., Zingaretti, P., 2008. Selection criteria of training set for optimal land cover discrimination in automatic segmentation. *28th EARSel Symposium, ISTANBUL*, 2-5 June 2008 "Remote Sensing for a Changing Europe".

THE REGISTRATION OF 3-D MODELS AND A 2-D IMAGE USING POINT AND LINE FEATURES

T. Teo [a, *], L. Chen [b]

[a] Dept. of Civil Engineering, National Chiao Tung University, Hsinchu, Taiwan 30010 - tateo@mail.nctu.edu.tw
[b] Center for Space and Remote Sensing Research, National Central University, Taoyuan, Taiwan 32001 - lcchen@csrsr.ncu.edu.tw

KEY WORDS: Image, Three-dimensional, Building, Registration, Orientation

ABSTRACT:

Space registration of 2-D images and 3-D models is an important task for geoinformation applications. Space registration connects the spatial relationship between the image space and object space. It can be implemented by using different control entities like control points, control lines, control surfaces, etc. 3-D models provide both point and line features. Hence, we establish a procedure to determine the image orientation by integrating these features. The objective of this investigation is to combine the point and linear features in space registration. The proposed scheme utilizes collinearity equations in determining the orientation. In this investigation, we compare three kinds of collinearity equations. The first one is a point-based formulation. The second one is line-based equations. The third one is a collinearity model that combines the point and line features. The test data include a simulation data, an aerial image and a close-range image. The experimental results indicate that the proposed scheme is flexible and reliable.

1. INTRODUCTION

A crucial aspect of integrating different geoinformation data is to establish a common reference frame (Zitova and Flusser, 2003). Registration of image and vector data is an important task for various applications, such as cartography, texture mapping, GIS technology, and others. Space registration establishes the spatial relationship between the image space and object space. It can be implemented by using different control entities like points, lines, surfaces, etc. The objective of this investigation is to combine the point and linear features in space registration.

The major work of space registration is to determine the exterior orientation parameters of image data (also called space resection or triangulation). There are three kinds of equations in determining the exterior orientations, i.e., collinearity, coplanarity and coangularity equations (Grussenmeyer and Al Khalil, 2002). The collinearity equations are well-known for orientation determination in photogrammetry field. These equations describe the collinearity geometry of perspective center, image point and object point. They use the intersection of linear-rays to determine the exterior orientations. On the other hand, the coplanarity equations describe the coplanarity geometry of a perspective center, a line in the image space and respective lines in the object space. It uses the intersection of planes to determine the exterior orientations. For coangularity equations, they indicate the coangularity condition among the angles of a perspective center and two object points and the respective angles in a camera frame.

The control entities are used to solve the equations in the orientation modeling. The entities include control points (Wolf and Dewitt, 2000), control lines (Akav et al., 2004; Habib et al., 2005; Jung and Boldo, 2004), control surfaces (Jaw, 2000), control patches (Jaw and Wu, 2006), etc. The control points represent a set of 2-D point features in the image space and 3-D point features in the object space. The point feature is easy to implement when comparing to others. The control line is a set of 2-D line features in the image space and 3-D line features in the object space. This kind of line features mainly occur in man-made objects like buildings. The control surface describes a set of 2-D point features in the image space and respective 3-D surface models. The control patch includes an image chip database which is used to define the location of the control points. The control patch usually uses image chip database to improve the automation of the point measurement.

The point-based triangulation is widely used in photogrammetric softwares as the control point is the basic control feature. It can be extended to other entities. In some typical scenarios, the linear feature can be measured more flexibly than point feature. Hence, the linear features are often selected in triangulation besides the point features. The vector data provide both control point and control line features. Several investigations have been reported on point-based or linear-based space resection (Karjalainen et al., 2006) from vector data. However, there is a lack of investigation to combine the point and linear features simultaneously on space resection.

In this investigation, we establish a procedure of image orientation determination by integrating point and linear features. The proposed scheme utilizes collinearity equations in the orientation determination. In this paper, we compare three kinds of collinearity equations. The first one is a point-based formulation. The second one is line-based equations. The third one is a joint equation which combines the point and linear features. The test data include a simulation data, an aerial image and a close-range image.

* Corresponding author.

2. METHODOLOGIES

Space registration utilizes space resection to obtain the exterior orientation parameters of image. We introduce three different models in this paper, i.e., point-based space resection, line-based space resection and joint model.

2.1 Point-based Space Resection

The point-based triangulation employs the collinearity equations to solve the exterior parameters. The collinearity equations define the condition of a perspective center, a point in image space and its corresponding object points that are on a straight line. Collinearity equations are shown in Equation 1. Figure 1 illustrates the collinearity condition. As the equations are non-linear with respect to the parameter, we need to linearize the equations and solve the parameters iteratively. The vector, matrix and the elements are shown in Equations 2 and 3. A more detailed description of the procedure can be found in (Wolf and Dewitt, 2000).

$$x = x_0 - f\frac{m_{11}(X-X^c)+m_{12}(Y-Y^c)+m_{13}(Z-Z^c)}{m_{31}(X-X^c)+m_{32}(Y-Y^c)+m_{33}(Z-Z^c)}$$
$$y = y_0 - f\frac{m_{21}(X-X^c)+m_{22}(Y-Y^c)+m_{23}(Z-Z^c)}{m_{31}(X-X^c)+m_{32}(Y-Y^c)+m_{33}(Z-Z^c)} \quad (1)$$

In Equation 1, (x, y) are the image coordinates; (x_0, y_0) are the principal points; (X, Y, Z) are the object coordinates; (X^c, Y^c, Z^c) are the coordinates of perspective center; $(m_{11}\sim m_{33})$ are the elements of rotation matrix from rotation angles (ω, φ, κ); f is the focal length.

$$V_p = B_p\Delta_p - L_p \quad (2)$$

In Equation 2, V_p is the vector of residual errors; B_p is the matrix of the coefficients of unknowns; Δ_p is the vector of unknown corrections of exterior parameters; L_p is the vector of constant terms. In Equation 3, F and G are the observation functions from collinearity equations.

$$\begin{bmatrix} V_{x1} \\ V_{y1} \\ \cdot \\ \cdot \\ V_{xn} \\ V_{yn} \end{bmatrix} = \begin{bmatrix} \frac{\partial F_1}{\partial X^c} & \frac{\partial F_1}{\partial Y^c} & \frac{\partial F_1}{\partial Z^c} & \frac{\partial F_1}{\partial \omega} & \frac{\partial F_1}{\partial \varphi} & \frac{\partial F_1}{\partial \kappa} \\ \frac{\partial G_1}{\partial X^c} & \frac{\partial G_1}{\partial Y^c} & \frac{\partial G_1}{\partial Z^c} & \frac{\partial G_1}{\partial \omega} & \frac{\partial G_1}{\partial \varphi} & \frac{\partial G_1}{\partial \kappa} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \frac{\partial F_n}{\partial X^c} & \frac{\partial F_n}{\partial Y^c} & \frac{\partial F_n}{\partial Z^c} & \frac{\partial F_n}{\partial \omega} & \frac{\partial F_n}{\partial \varphi} & \frac{\partial F_n}{\partial \kappa} \\ \frac{\partial G_n}{\partial X^c} & \frac{\partial G_n}{\partial Y^c} & \frac{\partial G_n}{\partial Z^c} & \frac{\partial G_n}{\partial \omega} & \frac{\partial G_n}{\partial \varphi} & \frac{\partial G_n}{\partial \kappa} \end{bmatrix} \begin{bmatrix} \Delta X^c \\ \Delta Y^c \\ \Delta Z^c \\ \Delta\omega \\ \Delta\varphi \\ \Delta\kappa \end{bmatrix} - \begin{bmatrix} F_1 \\ G_1 \\ \cdot \\ \cdot \\ F_n \\ G_n \end{bmatrix} \quad (3)$$

2.2 Line-based Space Resection

Line-based space resection applies the linear features as the control features. The linear features can be a straight line or a high-order line (Habib et al., 2003a). The straight lines are mostly selected in line-based space resection. The mathematic model of space resection includes collinearity and coplanarity conditions. Since the collinearity condition is geometrically stronger than the coplanarity one (Schenk, 2004), we select the collinearity approach in this study. Figure 2 illustrates space resection using linear features.

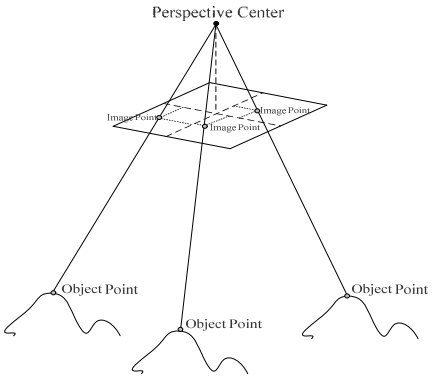


Figure 1. Illustration of point-based triangulation.

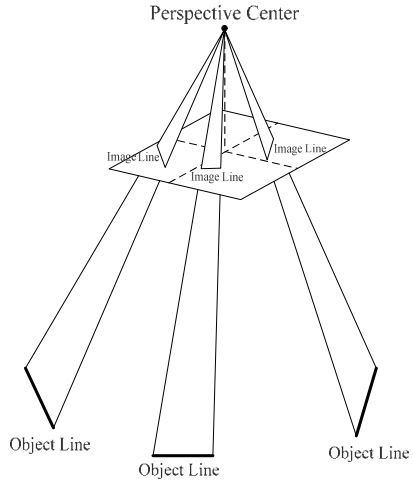


Figure 2. Illustration of line-based triangulation.

First, we need to establish a mathematic model for the line segment. The easiest way to represent a line segment is to use two end points. We can calculate the line parameter from these two end points. The line parameters include the starting point and direction vector. The mathematic model of a line can be formulated as Equation 4. The parameters are not independent and they will cause the ill-condition in solving the exterior orientations. In order to solve the ill-posed problem, we compute the intersection point of the line and plane of Z=0 first. Then, we select the starting point which locates in the plane. The direction vector is also normalized to reduce the parameters. Finally, the line parameters are reduced from six to four (Ayache and Faugeras, 1989). The parameters reduction can be referred to plane X=0, Y=0 or Z=0. The selection of reference planes is based on the angle between line segment and normal vector of plane. The smaller angle represents the better geometry. Hence, we can avoid the problem of no intersection point between line segment and reference when they are parallel. Equation 4 demonstrates the function when reference plane Z=0 is applied. Figure 3 shows the geometry of the line segment.

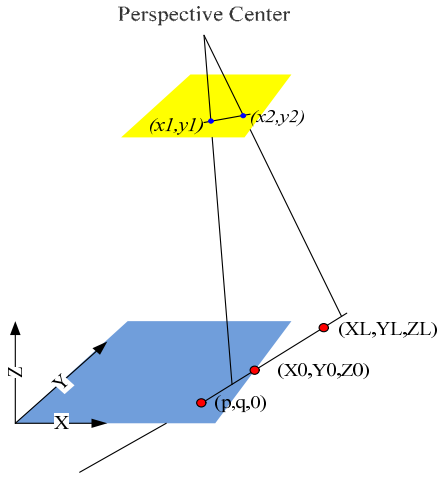


Figure 3. Representation of line-segment.

$$\begin{bmatrix} X_s \\ Y_s \\ Z_s \end{bmatrix} = \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix} + t\begin{bmatrix} dX \\ dY \\ dZ \end{bmatrix} = \begin{bmatrix} p \\ q \\ 0 \end{bmatrix} + s\begin{bmatrix} a \\ b \\ 1 \end{bmatrix} \quad (4)$$

In Equation 4, (Xs, Ys, Zs) are the coordinates of a point located in a line; (X_0, Y_0, Z_0) are the starting point of a line; (dX, dY, dZ) are the direction vector; (p, q, 0) are the starting point through the plane of Z=0; (a, b, 1) are the normalized direction vector; t and s are the scale factor of a line.

We establish the line-based collinearity equations by combining Equations 1 and 4. The new equation is shown as Equation 5. Compared with the point-based collinearity equations, the additional unknowns for line-based equations are the scale factor of a line. We linearize the equations and solve the parameters iteratively. The matrix and the elements are shown in Equations 6 and 7.

$$x = x_0 - f\frac{m_{11}(p+s*a-X^c)+m_{12}(q+s*b-Y^c)+m_{13}(s-Z^c)}{m_{31}(p+s*a-X^c)+m_{32}(q+s*b-Y^c)+m_{33}(s-Z^c)}$$
$$y = y_o - f\frac{m_{21}(p+s*a-X^c)+m_{22}(q+s*b-Y^c)+m_{23}(s-Z^c)}{m_{31}(p+s*a-X^c)+m_{32}(q+s*b-Y^c)+m_{33}(s-Z^c)} \quad (5)$$

In Equation 5, (x, y) are the image coordinates; (x_0, y_0) are the principal points; (X, Y, Z) are the object coordinates; (p, q, 0) and (a, b, 1) are the parameters of control line; s is the scale factor of a line; (m_{11}~m_{33}) are the elements of rotation matrix from rotation angles (ω, φ, κ); f is the focal length.

$$V_l = \begin{bmatrix} B_{l1} & B_{l2} \end{bmatrix}\begin{bmatrix} \Delta_p \\ \Delta_l \end{bmatrix} - L_l \quad (6)$$

In Equation 6, V_l is the vector of residual errors; B_{l1} and B_{l2} are the matrix of the coefficients of unknowns; Δ_p is the vector of unknown corrections of exterior parameters; Δ_l is the vector of unknown corrections of scale factor; L_l is the vector of constant terms.

$$\begin{bmatrix} V_{x1} \\ V_{y1} \\ \cdot \\ \cdot \\ V_{xn} \\ V_{yn} \end{bmatrix} = \begin{bmatrix} \frac{\partial F_1}{\partial X^c} & \frac{\partial F_1}{\partial Y^c} & \frac{\partial F_1}{\partial Z^c} & \frac{\partial F_1}{\partial \omega} & \frac{\partial F_1}{\partial \varphi} & \frac{\partial F_1}{\partial \kappa} & \frac{\partial F_1}{\partial s_1} & \cdot & 0 \\ \frac{\partial G_1}{\partial X^c} & \frac{\partial G_1}{\partial Y^c} & \frac{\partial G_1}{\partial Z^c} & \frac{\partial G_1}{\partial \omega} & \frac{\partial G_1}{\partial \varphi} & \frac{\partial G_1}{\partial \kappa} & \frac{\partial G_1}{\partial s_1} & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \frac{\partial F_n}{\partial X^c} & \frac{\partial F_n}{\partial Y^c} & \frac{\partial F_n}{\partial Z^c} & \frac{\partial F_n}{\partial \omega} & \frac{\partial F_n}{\partial \varphi} & \frac{\partial F_n}{\partial \kappa} & \cdot & \cdot & \frac{\partial F_n}{\partial s_o} \\ \frac{\partial G_n}{\partial X^c} & \frac{\partial G_n}{\partial Y^c} & \frac{\partial G_n}{\partial Z^c} & \frac{\partial G_n}{\partial \omega} & \frac{\partial G_n}{\partial \varphi} & \frac{\partial G_n}{\partial \kappa} & \cdot & \cdot & \frac{\partial G_n}{\partial s_o} \end{bmatrix}\begin{bmatrix} \Delta X^c \\ \Delta Y^c \\ \Delta Z^c \\ \Delta\omega \\ \Delta\varphi \\ \Delta\kappa \\ \Delta s_1 \\ \cdot \\ \Delta s_o \end{bmatrix} - \begin{bmatrix} F_1 \\ G_1 \\ \cdot \\ \cdot \\ F_n \\ G_n \end{bmatrix} \quad (7)$$

2.3 Space Resection by Point and Line Features

In this section, we introduce space resection by integrating the point and linear features. Figure 4 shows the idea of space resection using point and linear features. There are some advantages of using both features in space resection. For example, when the point and linear features are both available, we can use the complete information to determine the exterior orientations simultaneously. Moreover, the controlling capability of control points is geometrically stronger than control line. If we add some control points with control line in space resection, it improves the accuracy of exterior orientations. We establish a joint model by combining Equations 2 and 6. The new equation is shown as Equation 8.

This join adjustment model needs to linearize for least squares adjustment. A number of initial values are needed to obtain an approximate value. The initial values of exterior orientation parameters can be determine by Direct Linear Transformation (Kobayashi and Mori, 1997). The other initial value is the scale factor of a line segment which can be derived from two end points. Nowadays, GPS is mounted on specific camera for camera position. This information is useful for the initial values of a camera position.

$$\begin{bmatrix} V_p \\ V_l \end{bmatrix} = \begin{bmatrix} B_p & 0 \\ B_{l1} & B_{l2} \end{bmatrix}\begin{bmatrix} \Delta_p \\ \Delta_l \end{bmatrix} - \begin{bmatrix} L_p \\ L_l \end{bmatrix} \quad (8)$$

In Equation 8, V_p and V_l are the matrix of residual errors; B_p, B_{l1} and B_{l2} are the matrix of the coefficients of unknowns; Δ_p is the matrix of unknown corrections of exterior parameters; Δ_l is the matrix of unknown corrections of scale factor; L_p and L_l are the matrix of constant terms.

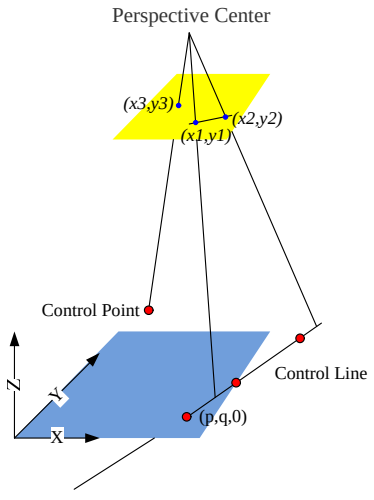


Figure 4. Space resection by the integration of control point and control line.

3. EXPERIMENTAL RESULTS

Three data sets are selected in the experiment. The first one is a simulation data. The second one is an aerial image. The third one is a close-range image.

3.1 Simulation Data

Simulation data is used to verify the procedure. We simulate an aerial image with vertical imaging. In this simulation, we add random errors to the control features. The random errors are 0.5 pixels in the image space and 0.5 meters in the object space. We simulate 38 control points (CP) and 41 control lines (CL) in this study. The 372 independent check points (ICP) are used to evaluate the accuracy of the exterior parameters. Figure 5 shows the distribution of simulation data. For point-based space resection, the root mean square errors (RMSE) of ICP are 0.62 and 0.88 pixels in two directions. For line-based space resection, though the number of control line is more than the number of control point, the RMSE of ICP is still 0.87 and 1.01 pixels in *x* and *y* directions. The result of line-based space resection is not as good as point-based space resection. We combine 3 control points and 41 control lines in our proposed method. The RMSE of ICP improves to 0.79 and 1.00 pixels. We also combine all the control points and control lines in space resection. The RMSE of ICP improves to 0.38 and 0.56 pixels. Tables 1 and 2 summarize the results of simulation data without and with random error.

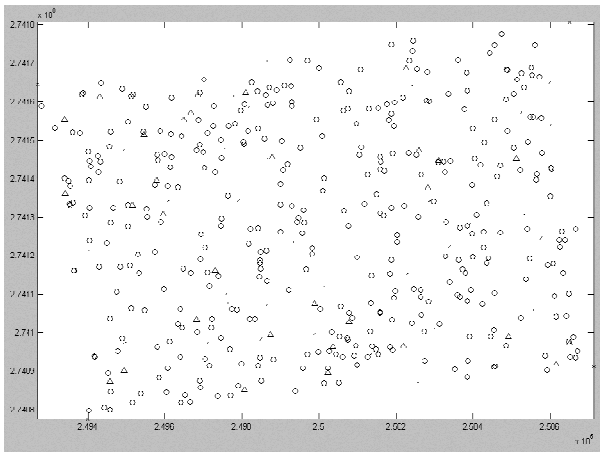


Figure 5. Distribution of control features of aerial image.

Table 6. Results of simulation data (without random error).

Unit: pixel	Point	Line	Point+Line
No. CP	38	0	38
No. CL	0	41	41
No. ICP	372	372	372
Mean Sample	-8.93E-06	-9.24E-06	-9.27E-06
Mean Line	3.12E-06	1.74E-06	1.84E-06
RMSE Sample	2.24E-04	2.26E-04	2.29E-04
RMSE Line	3.33E-04	3.34E-04	3.39E-04

Table 7. Results of simulation data (with random error).

Unit: pixel	Point	Line	Point+Line		
No. CP	38	0	1	3	38
No. CL	0	41	41	41	41
No. ICP	372	372	372	372	372
Mean Sample	0.44	0.44	0.41	0.27	0.03
Mean Line	0.68	0.08	0.06	-0.05	0.37
RMSE Sample	0.62	0.87	0.86	0.79	0.38
RMSE Line	0.88	1.01	1.00	1.00	0.56
σ ω	0.0005	0.0005	0.0005	0.0005	0.0004
σ φ	0.0004	0.0005	0.0005	0.0005	0.0003
σ κ	0.0001	0.0001	0.0001	0.0001	0.0001
σ Xc	0.5045	0.6916	0.6835	0.6815	0.3845
σ Yc	0.6903	0.6939	0.6857	0.6809	0.4859
σ Zc	0.1596	0.1871	0.1842	0.1840	0.1142

3.2 Aerial Image

The second data is an aerial image with 3-D building models. The image was acquired by an UltraCamD with 1/12,000 image scale. The image size is 11500 x 7500 pixels. The ground resolution of the image is about 12cm. The accuracies of 3-D building corners are 20cm and 35cm in horizontal and vertical direction. We measure 10 control points and 10 control lines from the image and maps. The number of independent check points is 25. Figure 8 shows the distribution of control features superimposed with aerial image. The corresponding 3-D building models are shown as Figure 9. The geometric characteristics of aerial image are similar to simulation data. The RMSE of ICP is 1.87 and 2.26 pixels in *x* and *y* directions when we apply the point-based space resection. The RMSE of line-based space resection is 2.50 and 2.86 pixels. The RMSE of proposed method is 2.17 and 2.54 pixels. Table 10 summarizes the results of aerial image.

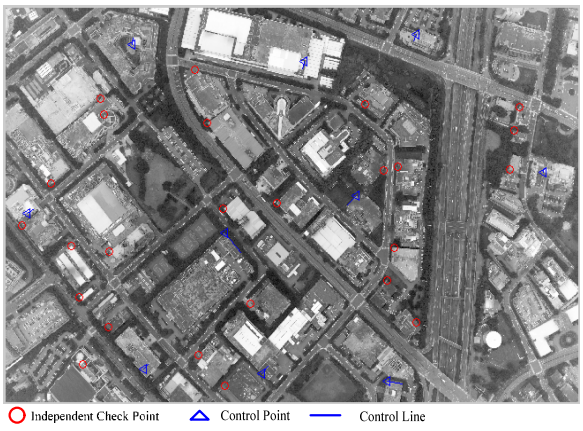


Figure 8. Distribution of control features of aerial image.

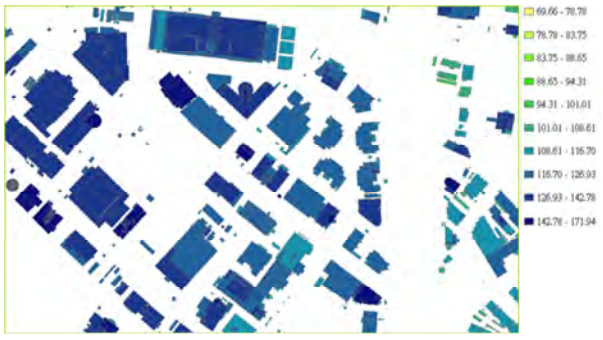


Figure 9. 3-D models for aerial image.

Table 10. Results of aerial image.

Unit: pixel	Point	Line	Point+Line		
No. CP	10	0	1	10	10
No. CL	0	10	10	1	10
No. ICP	25	25	25	25	25
Mean Sample	0.86	1.05	1.36	0.86	1.09
Mean Line	-0.92	-1.77	-1.67	-0.89	-1.35
RMSE Sample	1.87	2.49	2.46	1.87	2.17
RMSE Line	2.26	2.86	2.79	2.24	2.54

3.3 Close-range Image

For close-range image, the image scale is about 1/1,700. The resolution of the image is about 1cm. The image size is 4288 x 2848 pixels. The rough camera position is recorded by a GPS, which is mounted on the camera. The lens distortions are pre-calibrated by using commercial software, i.e., PhotoModeler. As the building detail in a building model is not very high, we can only measure a few control features. The accuracy of reference coordinates are 20cm and 35cm in horizontal and vertical direction.. The number of control points and lines are 4 and 7 respectively. We also measure 3 independent check points for accuracy evaluation. The distribution of control features are shown in Figure 11. The respective building models are shown in Figure 12. Notice that, the length of the control lines is quite long related to the image frame. For point-based space resection, the RMSE of ICP is 14.44 and 9.20 pixels in *x* and *y* directions. The RMSE of line-based space resection is 16.75 and 6.13 pixels. We also combine all the control points and control lines in space resection. The RMSE of ICP improves to 15.56 and 6.45 pixels. Table 10 summarizes the results of close-range image.

Figure 11. Distribution of control features of ground image.

Figure 12. 3-D models for close-range image.

Table 13. Results of close-range image.

Unit: pixel	Point	Line	Point+Line
No. CP	4	0	4
No. CL	0	7	7
No. ICP	3	3	3
Mean Sample	5.82	9.53	7.76
Mean Line	0.42	-1.17	0.17
RMSE Sample	14.44	16.75	15.56
RMSE Line	9.20	6.13	6.45

4. CONCLUSIONS

In this research, we have proposed a feasible scheme to obtain the exterior orientations that integrates point and linear features. The simulation data show that the result of point-based adjustment is better than the line-based adjustment. The combine adjustment of these two features may improve the accuracy. If all the available data involve the adjustment, the best result is expected when compare to the other models. The accuracy of orientation for aerial image is around 2.5 pixels in this investigation. The resolution of image is 12cm and the accuracy of reference data is about 20cm to 35cm. For close-range image, the accuracy of orientations is better than 15 pixels. The resolution of image is 1cm. Hence, the accuracy in the object space is about 15cm. The results are based on the accuracy of 3-D building models. Thus, if the quality of model improves, the higher accuracy is expected.

REFERENCES

Ayache, N., and Faugeras, O.D. 1989. *Maintaining representations of the environment of a mobile robot*, IEEE transactions on robotics and automation, 5(6), pp. 804-819.

Akav, A., Zalmanson, G.H. and Doytsher, Y., 2004. *Linear feature based aerial triangulation, International Archives of Photogrammetry*, Remote Sensing and Spatial Information Sciences, pp. 7-12.

Grussenmeyer, P. and Al Khalil, O., 2002. Solutions for exterior orientation in photogrammetry: a review. Photogrammetric Record, 17(100), pp. 615-634.

Habib, A., Ghanma, M., Morgan, M. and Al-Ruzouq, R., 2005. *Photogrammetric and lidar data registration using linear features*. Photogrammetric Engineering and Remote Sensing, 71(6) ,pp. 699-707.

Habib, A., Lin, H.T. and Morgan, M., 2003a. *Autonomous space resection using point- and line-based representation of free-form control linear features.* Photogrammetric Record, 18(103), pp. 244-258.

Jaw, J.J., 2000. *Control surface in aerial triangulation, International Archives of Photogrammetry*, Remote Sensing and Spatial Information Sciences, pp. 444-451.

Jaw, J.J. and Wu, Y.S., 2006. Control patches for automatic single photo orientation. Photogrammetric Engineering And Remote Sensing, 72(2), pp. 151-157.

Jung, F. and Boldo, d., 2004. *Bundle adjustment and incidence of linear features on the accuracy of external calibration parameters*, International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, pp. 19-24.

Karjalainen, M., Hyyppa, J. and Kuittinen, R., 2006. *Determination of exterior orientation using linear features from vector maps.* Photogrammetric Record, 21(116), pp. 329-341.

Kobayashi, K. and Mori, C., 1997. *Relations between the coefficients in the projective transformation equations and the orientation elements of a photograph.* Photogrammetric Engineering & Remote Sensing, 63(9), pp. 1121–1127.

Schenk, T., 2004. *From point-based to feature-based aerial triangulation*. ISPRS Journal of Photogrammetry and Remote Sensing, 58(5-6), pp. 315-329.

Wolf, P.R. and Dewitt, B.A., 2000. Elements of Photogrammetry: With Applications in GIS. McGraw-Hill, 608 pp.

Zitova, B. and Flusser, J., 2003. *Image registration methods: a survey*. Image and Vision Computing, 21(11), pp. 977-1000.

ACKNOWLEDGEMENTS

This investigation was partially supported by the National Science Council of Taiwan under project number NSC 98-2218-E-009 -018. The authors would like to thank the Chinese Society of Photogrammetry and Remote Sensing of Taiwan for providing the test data sets.

ESTIMATION OF AEROSOL AND AIR QUALITY FIELDS WITH PM MAPPER – AN OPTICAL MULTISPECTRAL DATA PROCESSING PACKAGE

T. N. Thanh Nguyen [a, b, *], S. Mantovani [b, c], M. Bottoni [c]

[a] University of Ferrara, Department of Mathematics, Via Saragat 1, I-44122, Ferrara, Italy - thanhntn@meeo.it
[b] MEEO S.r.l., Via Saragat 9, I-44122, Ferrara, Italy - mantovani@meeo.it
[c] SISTEMA GmbH, Währingerstrasse 61, A-1090, Wien, Austria - bottoni@sistema.at

KEY WORDS: Atmosphere, pollution, monitoring, retrieval, system, multi-spectral, global.

ABSTRACT:

In the frame of the Remote Sensing applications applied to MODIS data collected by the polar orbiting satellites Terra and Aqua, operated by the NASA, we present PM MAPPER, a novel data processing system developed for air pollution monitoring. Our system has derived from the MODIS data an updated set of information consisting of AOT, $PM_{2.5/10}$, AQI, and surface information with increased spatial resolutions up to 3x3 km^2. With the fine spatial resolution and augmented background information, the software is effective in monitoring air pollution at local scale, especially over small urban areas with complicated topography. We carried out a validation on the data set covering Italy in a period of six months to evaluate the system's performance. The validation outcomes show that our results have good quality in comparison with MODIS standard products and a higher capacity in retrieving AOT information over land areas, especially coastlines where are nearly empty in the MODIS products. Besides, integrated surface information could be useful for further improvements of aerosol derivation.

1. INTRODUCTION

Since the launch in 1999 and 2002, respectively, of the two Earth Observing System's (EOS), polar orbiting satellites Terra and Aqua, operated by the National Aeronautic and Space Administration (NASA), a large amount of remote sensing data has been made available to the scientific community. These data, collected by the MODerate resolution Imaging Spectrometer (MODIS), described in detail in Salomonson et al., 1989, cover a spectral range from 0.412 µm to 14.2 µm, partitioned in 36 channels and are characterized by high spatial resolution, up to 250 m at nadir. Original MODIS data are pre-elaborated by a software package, the most recent version of which is known as "Collection 005" and is described in detail in Remer et al., 2004. One of the most important products of the MODIS Atmosphere algorithms applied in Collection 005 is, among other scientific information, the retrieval of the Aerosol Optical Thickness (AOT), representative of the amount of particulates present in a vertical column of the Earth's atmosphere. As reported in the above mentioned reference, these data have been used in recent years for a large amount of scientific and technological endeavours, ranging from research on climate evolution, due to aerosol impact on the characteristics of the atmosphere, to dust sedimentation in the depth of the oceans.

The use of MODIS data for air pollution monitoring, described in detail in Wang et al (2003), has recently become a promising approach because it provides global distribution of aerosol and its properties for deriving indirect estimates of particulate matter concentration, one of the major pollutants that affect air quality. The best available resolution provided by MODIS standard aerosol products, up to now, is 10x10 km^2 that is adequate for monitoring at global scale but not fine enough at local scale, such as urban areas. Several researchers have been aiming at deriving from the original MODIS information, presented as aerosol data integrated over 10x10 km^2, more detailed information covering areas of 1.5x1.5 km^2 (Oo et al., 2008) or 1x1 km^2 (Li et al., 2005). The latter reference shows how high resolution remote sensing can be applied to densely populated urban areas, like the Hong Kong metropolitan region, to infer about pollutant distributions and air quality index. PM MAPPER developed by our research group can be set on the wake of the above mentioned works because it aims at deriving from the original data an updated set of information including Aerosol Optical Thickness (AOT), Particulate Matter concentration ($PM_{2.5}$, PM_{10}), and Air Quality Index (AQI) with increased spatial resolutions to 3x3 km^2, taking advantage of new criteria for deriving from the MODIS data the spectral characteristics of land areas classified in a series of up to 40 different classes. In Section 2 of this article, we present the strategy and methodology applied to analyse the MODIS data in comparison with the standard one. Results obtained with our version of the data analysis package will be presented in Section 3, while validation of our results, in comparison with the MODIS standard products, will be discussed in Section 4. On the basis of the results obtained and of the validations made thus far, some conclusions will be drawn in section 5, together with hints about necessary future developments of the new algorithms.

2. METHODOLOGY

The methodology for deriving AOT information from MODIS data consists of land and ocean algorithms following the Look-Up Table (LUT) approach which matches satellite observations to simulated values in LUT to derive aerosol concentration and its properties. The algorithms have performed on land/water pixels which are cloud-free. The covered region (land/water) is determined by pixel geographical information, and clouds are

* Corresponding author

rubbed out by cloud screening process described as MODIS cloud mask in Ackerman et al (1998).

The algorithm over ocean uses observed radiances in seven wavelengths (0.47-2.13 μm) as input data and assumes a bi-modal log-normal distribution for the size of aerosol particles which are then classified in 5 small and 6 large groups referred as "aerosol modes". In the retrieval process, for each combination of two aerosol modes with size ration η, the total radiance at band 0.55 μm is calculated in order to find the best ratio η and the particle sizes of the small and large modes that give the best fit to the measurements in this band. The selection of aerosol model is performed until the residual error computed over 6 channels (0.55-2.13 μm) is minimized. During this process, the optical thickness is derived in all seven channels.
The algorithm over land uses Dense Dark Vegetation (DDV) approach which derives AOT over dark pixels at the blue and red channels (0.47 and 0.66 μm). The dark pixels, corresponding to vegetation or other land regions, are identified by using correlations between mid-IR reflectance, usually at 2.1 μm or 3.8 μm, and blue and red channels. However, to avoid the effects of residual cloud contamination and of some bright surfaces, 50% brightest and 20% darkest pixels are eliminated. The surface reflectances are computed from the mid-IR channel mentioned above, based on the empirical relationship between them as given in Kaufman et al., 1997. Based on the LUTs, the aerosol concentration and its properties are derived from the appropriate dynamical aerosol model (urban/industrial, biomass burning, or dust) selected by the process mentioned in detail in Kaufman and Tanré, 1997.

In an effort to improve the spatial resolution of the MODIS aerosol products from 10x10 km^2 to 3x3 km^2, derive surface information and then estimate PM concentration and AQI maps for air pollution monitoring, we developed a new software package called PM MAPPER (MEEO S.r.l., 2009a). Figure 1 outlines in a schematic way the differences and correlations between the original MODIS system and the new one in which our package uses a software program called SOIL MAPPER (MEEO S.r.l., 2009b) instead of MODIS cloud mask module, modifies the aerosol algorithms over land and ocean, and estimates PM concentration and AQI maps from AOT results.

SOIL MAPPER use reflectances in eight wavelengths (0.66, 0.87, 0.47, 0.55, 1.64, 2.13, 11.03, and 12.02 μm) to identify land, water, and cloudy pixels before applying aerosol retrieval procedure. This software has been developed in the past years, based upon spectral fuzzy rule-based per-pixel classification method, originally presented and discussed in Baraldi et al., 2006. Unlike the MODIS cloud mask module, it generates a Land Cover (LC) map by using a rule based system architecture consisting of a two-level processing steps. To deal with cloud in particular, the software follows the Automatic Cloud Cover Assessment (ACCA) approach (Irish, 2000) then enhanced by spectral signature techniques for discrimination of cloud and snow covered areas (Dozier, 1989). Using the novelty of this program package, we are able to determine a set of 57 different classes, out of which 40 refer to different land types, from dense vegetation to bare soil, and the remaining classes refer to water or ice or snow covered land areas.

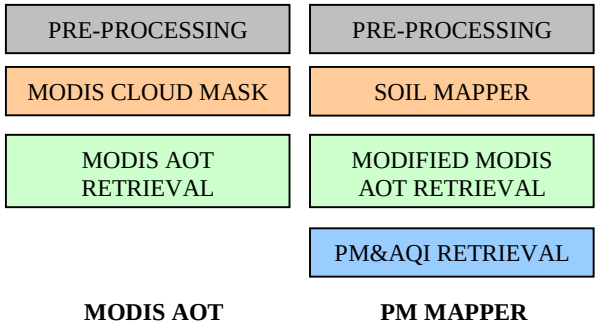


Figure 1. Main software modules of MODIS aerosol standard product and PM MAPPER software package.

In order to retrieve AOT concentration maps up to 3x3 km^2 spatial resolution, we modify both the original aerosol algorithms over land and ocean. All related procedures are considered and adapted on the smaller boxes sized 3x3 or 6x6 pixels of 1000 or 500 m MODIS measurements respectively instead of 10x10 or 20x20 pixels as presented in the original MODIS package. Processing on smaller areas leads to the requirement of adapting some thresholds, such as number of dark pixels required for a box in DDV algorithm. We solve all problems related to the definition of thresholds by the linear reduction method. Except for above modifications, the overall procedure follows the MODIS original method, as mentioned in the first part of this section.

PM&AQI retrieval is the last module added to estimate PM concentration map and then derive AQI map. PM is a complex mixture of solid and liquid particles that vary in size and composition, and can be divided into $PM_{2.5}$ or PM_{10} by theirs aerodynamic diameters. $PM_{2.5}$ and PM_{10} have been used to measure the air quality and can be estimated from AOT, based on the quantitative relationship between them as described in Wang et al., 2003 and Gupta et al., 2006. AQI and the related PM concentration afford useful information about the impact of air pollution on human health. This impact can be evaluated from the classification of air quality used in the US EPA and given in Table 2 (Environmental Protection Agency, 1999).

Index	Air Quality Category	$PM_{2.5}$ [μg/m^3]	PM_{10} [μg/m^3]
0-50	Good	0-15.0	0-50
51-100	Moderate	15.1-40.0	51-150
101-150	Unhealthy for sensitive groups	40.1-65.0	151-250
151-200	Unhealthy	65.1-150.0	251-350
201-300	Very unhealthy	150.1-250.0	351-420

Table 2. Air risk categories according to US EPA for $PM_{2.5}$ and PM_{10}.

Developed by using methodologies mentioned above, PM MAPPER package has required MODIS Level 1B HDF (Hierarchical Data Format) radiance and geolocation files as input and derived an updated set of information consisting of AOT, $PM_{2.5}$, PM_{10}, AQI, and surface classes at 3x3 km^2 spatial resolution. The system flowchart is shown in Figure , while the system's performance and results will be presented in more detail in the next section.

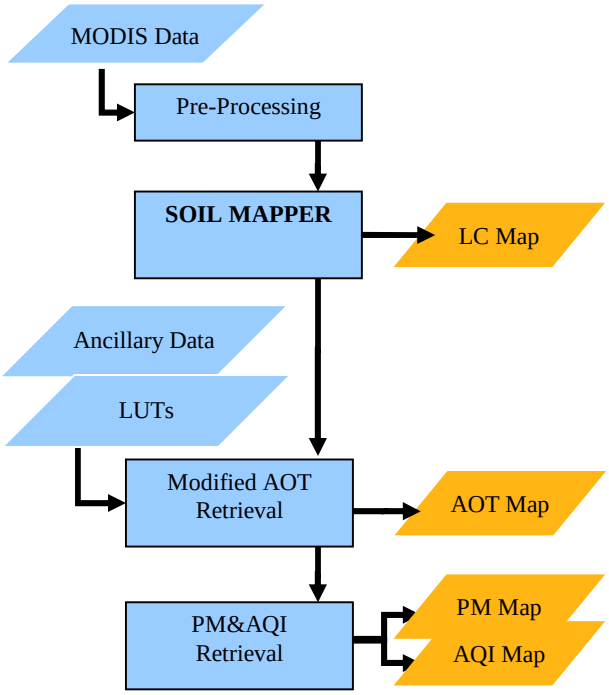


Figure 3. The system overview.

3. RESULTS

PM MAPPER package is developed on Linux operating system using C, Fortran, and IDL programming languages. One execution tested on a computer with Intel (R) Core(TM)2 Quad CPU Q9550 @2.83 GHz, 4Gb RAM and Ubuntu 9.10 platform, takes in average about 3 minutes and requires ~ 1GB of memory. In comparison with MODIS package performance (about 5 minutes and 200MB of memory), our software requires less time but more memory for an operation. The large memory requirement is justified by the more comprehensive results and by the higher resolution but also hints at the need of future software optimization.

The software outcome consists of Aerosol Optical Thickness (AOT), Land Cover (LC), Particulate Matter concentration ($PM_{2.5/10}$), and Air Quality Index (AQI) maps. Two salient characteristics of an output set are spatial resolution and land surface information.

The spatial resolution of AOT maps is increased to 3x3 km^2, which provides more detailed information about the distribution of AOT. In general, coarse spatial resolution, e.g. 10x10 km^2 of MODIS aerosol standard products, is suitable for pollution monitoring at global scale because of providing a general view. However, smaller locations with complicated topography, such as urban areas, require observation at local scale that implies more elaborate AOT information. One illustration of our products is shown in Figure 4. It covers the areas of Italy, Bosnia, Herzegovina, Croatia, the Mediterranean sea and the Adriatic sea. Although the considered range is at nation scale, this example for local air monitoring is appropriate from the viewpoint of spatial resolution. Our result, displayed in Figure 4 (b), in comparison with MODIS aerosol product in Figure 4 (a), guarantees the consistence of general spatial distribution of AOT, especially over seas. This, in theory, proves the stableness of ocean aerosol retrieval algorithm. However, differences can be observed over land (Italy, Bosnia, Herzegovina and Croatia), where our result presents a larger number of retrieval pixels. The coastline is another noticeable area filled by AOT values in our product and mostly empty in MODIS product (Figure 4(a)-(b)). This is due to the ineffectiveness of DDV algorithm over locations having both land and water pixels which, however, could be improved by applying processing data on smaller boxes as in our approach.

Together with the higher spatial resolution, the integrated surface information with AOT is another remarkable aspect of our results. For each pixel, corresponding to available AOT, the surface class is drawn, detailed in up to 57 values, out of which 40 are for land and 17 for other background (e.g. ice, snow, water, etc.). This integration provides another view of relationships between AOT and surfaces, which will be meaningful for investigation and correction of the factors affecting the AOT retrieval algorithm such as the correction of AOT bias on different land surfaces (Lary et al., 2009).

Starting from AOT maps, corresponding PM concentration and AQI maps are derived by our system as shown in Figure 5, which then is used directly for monitoring air pollution.

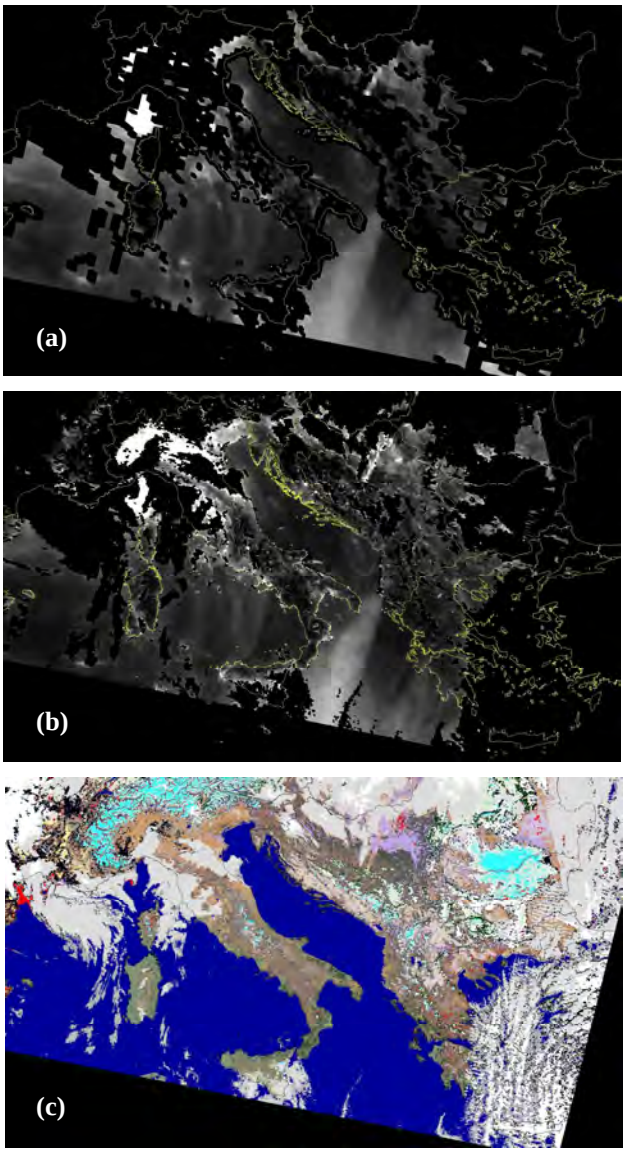


Figure 4. (a) MODIS AOT map at 10x10 km^2, (b) PM MAPPER AOT map at 3x3 km^2, and (c) classification map.

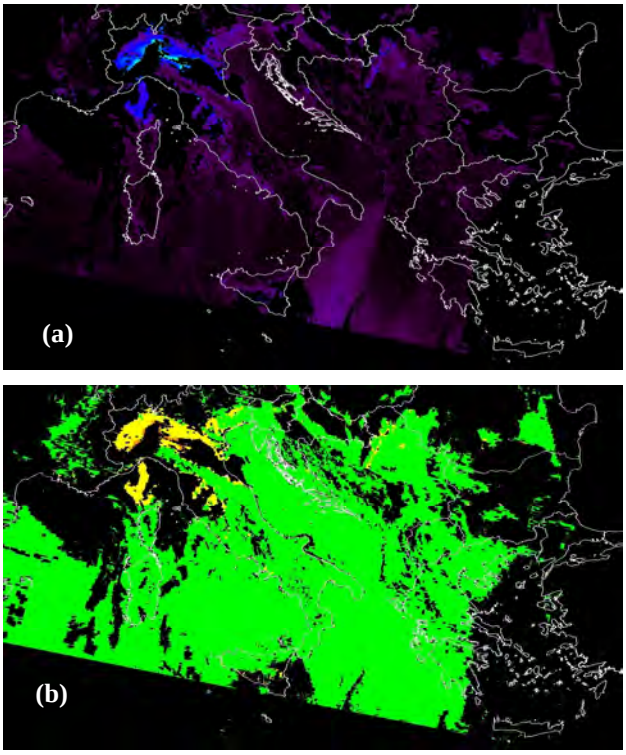


Figure 5. (a) PM concentration map and (b) AQI map derived from AOT map.

4. VALIDATION

In order to evaluate the PM MAPPER products, a twofold validation analysis has been performed. The first one consisted in assessing the quality of 3x3 km^2 spatial outputs. The second investigated the system performance more deeply over land areas by considering its results on different land surfaces. All assessments are based on the comparison of AOT values in our products and in MODIS aerosol products. In the validation, the data set consists of 15 images selected out of a total of 180 images covering Italy in a period of 6 months, from January 2008 to June 2008.

In the first validation, we evaluate the products at 3x3 km^2 in comparison with MODIS products at 10x10 km^2 over land, ocean, and both areas. Two parameters, correlation coefficient and the number of retrieved pixels, are considered. To avoid problems due to the different resolutions, all AOT maps are re-sampled at 1x1 km^2 resolution before any comparison. The results show the high correlation coefficients of our AOT with respect to MODIS AOT as plotted in Figure 6. The integrated AOT, a combination of aerosol over both land and ocean, has the correlation coefficient 0.88 in average with a deviation from 0.78 to 0.95. In more particular, the scattering plots of correlation coefficients between PM MAPPER AOT and corresponding MODIS AOT in the best and the worst cases are shown in Figure 7(a)-(b) respectively.

The number of retrieval pixels is representative for the map covering capacity. Figure 8(a) shows that the largest number of retrieved pixels is achieved by our software rather than by the original MODIS version (about 43.10% more pixels in average). This result is due to a large increase over land in spite of a small decrease over ocean (Figure 8 (b)-(c)). This improvement comes from our approach in which we applied AOT retrieval algorithm on smaller boxes and used SOIL MAPPER instead of MODIS cloud mask module to identify land, water, and cloudy pixels.

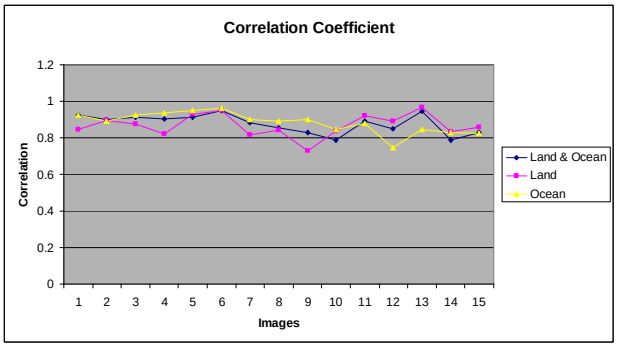


Figure 6. Correlation Coefficient of PM MAPPER with respect to MODIS version

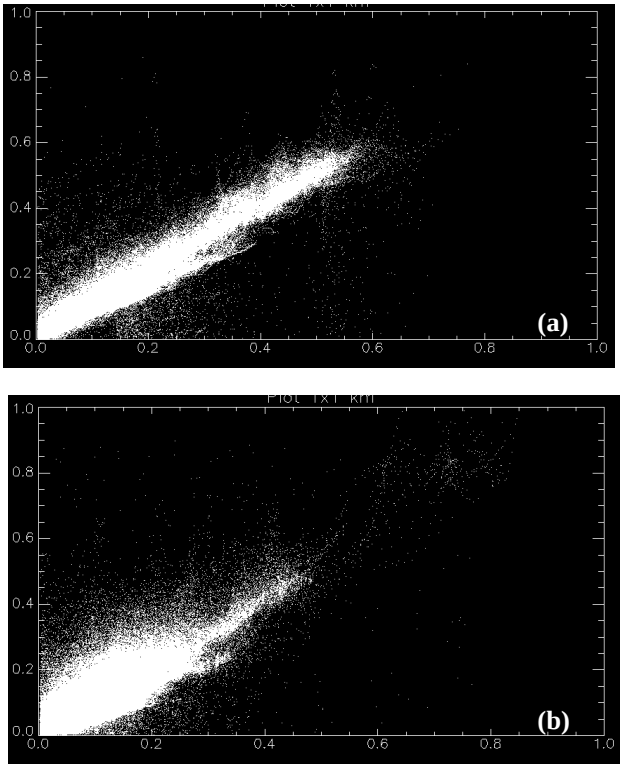
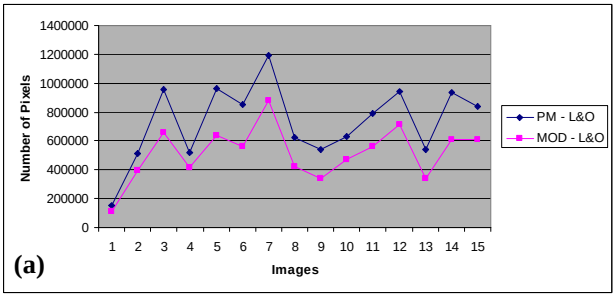


Figure 7. The scattering plot of (a) the best and (b) the worst correlation coefficient case, 0.95 and 0.78 respectively.



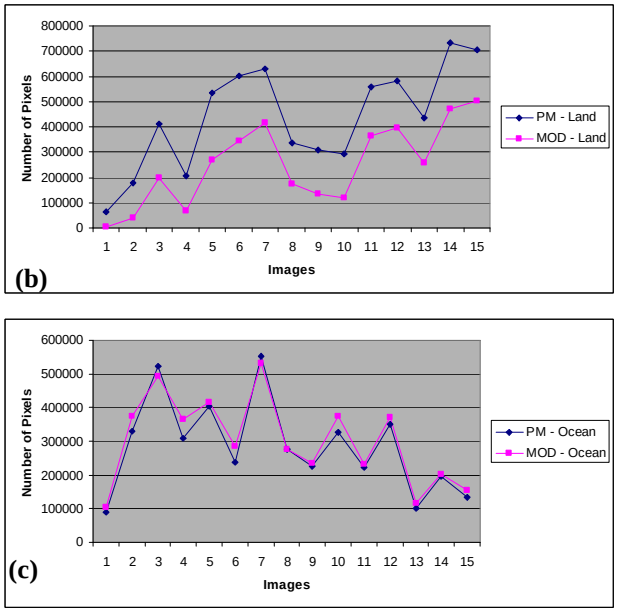


Figure 8. Number of retrieved AOT pixels on (a) Land and Ocean, (b) only Land, and (c) only Ocean.

The second validation is carried out to consider more deeply results at 3x3 km^2 on a stratified basis using integrated surface information detailed in up to 40 land classes (see Table 9). The purpose of the stratified analysis is to verify where the applicability of DDV algorithm is appropriate, that is, to investigate the relationship between the MODIS AOT values and PM MAPPER AOT values for different land cover classes.

For each land class, we calculated its darkness as the ratio of the number of dark pixels to the total number of pixels. One pixel is considered as a dark pixel when given conditions on the band 2.1 μm in DDV algorithm are satisfied. The darkness will decide the quality of aerosol optical thickness derivation over the class because aerosol derivation is affected by the uncertainty in the scattering phase function, as explained in Remer et al., 2004. The number of retrieved AOT pixels and the AOT correlation coefficient with respect to MODIS product are only computed over pixels having integrated information which is AOT and corresponding land label.

The validation results are divided into 4 groups as presented on Table 9. The first and second groups represent land classes having poor statistics on dataset and low correlation coefficients. In this case, more data need to be investigated. Group 3 includes two large and bright classes (16.3% and 46.1% of dark pixels for Bright Barren Land 2 and Bright Barren Land 4, respectively) where, in theory, the DDV retrieval algorithm is expected to be ineffective. However, the data analysis shows bad results only over Bright Barren Land 2 (correlation coefficient 0.627), whereas good results are obtained over Bright Barren Land 4 (correlation coefficient 0.809). The accuracy of dark pixel scanning over these classes can be considered as a reason for the noted differences. Even if this scanning performs well, over a very bright surface such as Bright Barren Land 2, the new improvements of aerosol derivation should be investigated. Group 4 consists of the residual classes that are dark and are characterized by a large number of retrieved pixels. Most of them have good correlation coefficients, but some can be candidates for improvements (i.e., Average Herbaceous Rangeland or Average Barren Land 2).

Labels	D.	N.	C.	G.
Bright Weak Vegetation	0	0	0.000	1
Bright Strong Shrub Rangeland	0	0	0.000	1
Dark Strong Shrub Rangeland	0	0	0.000	1
Strong Herbaceous Rangeland	0	0	0.000	1
Dark Barren Land 2	0	0	0.000	1
Bright Barren Land 1	0.989	409	0.387	2
Average Barren Land 1	1	1,450	0.539	2
Dark Barren Land 1	1	102	0.557	2
Strong Barren Land 2	0.982	2,775	0.674	2
Dark Barren Land 4	1	191	0.708	2
Dark Peat Bogs	1	99	0.709	2
Strong Barren Land 1	1	485	0.736	2
Bright Strong Vegetation	1	76	0.755	2
Dark Weak Vegetation	0.989	110	0.835	2
Bright Barren Land 2	0.163	105,765	0.627	3
Bright Barren Land 4	0.431	345,917	0.809	3
Average Herbaceous Rangeland	1	17,985	0.746	4
Average Barren Land 2	0.999	6,436	0.747	4
Dark Range Land	1	83,288	0.764	4
Dark Strong Vegetation	1	87,393	0.765	4
Bright Peat Bogs	0.998	3,806	0.782	4
Wet land	0.996	13,829	0.786	4
Shadow Barren Land	0.978	8,914	0.799	4
Dark Average Shrub Rangeland	1	50,557	0.809	4
Strong Barren Land 3	1	52,505	0.814	4
Shadow Vegetation	0.921	76,124	0.816	4
Bright Average Shrub Rangeland	0.999	697,678	0.816	4
Strong Barren Land 4	0.999	190,492	0.817	4
Dark Barren Land 3	1	129,072	0.828	4
Strong Barren Land 5	0.997	48,374	0.831	4
Bright Barren Land 3	0.908	41,569	0.833	4
Very Bright Average Vegetation 2	0.998	976,442	0.838	4
Average Barren Land 4	1	206,557	0.839	4
Mid tone Peat Bogs	1	5,225	0.841	4
Bright Average Vegetation	1	1,133,162	0.855	4
Very Bright Average Vegetation 1	1	392,399	0.856	4
Bright Rangeland	0.999	231,669	0.856	4
Dark Average Vegetation	0.999	308,450	0.863	4
Mid tone Rangeland	1	386,860	0.878	4
Average Barren Land 3	1	326,197	0.880	4

Table 9. The performance of PM MAPPER at 3x3 km^2 over difference land classes (D: Darkness, N: Number of retrieval AOT pixels, C: Correlation Coefficient, G: Group)

In conclusion, the first validation results show that PM MAPPER products have good quality in comparison with MODIS standard aerosol products and higher capacity in retrieving AOT information (larger number of retrieved pixels). The second validation reinforces the previous result with the high correlation coefficient over most of dense surface classes. This validation also groups and points out some potential land

classes where the aerosol retrieval algorithm needs to be improved.

5. CONCLUSION AND FUTURE WORKS

In this article, we presented the PM MAPPER, a multispectral data processing package for aerosol and air quality field estimation. The software has derived from MODIS data a set of information consisting of AOT, $PM_{2.5}$, PM_{10}, AOT, and surface classes with increased spatial resolutions up to 3x3 km^2. PM MAPPER products are characterized by fine spatial resolution and integrated background information, provided by the SOIL MAPPER, in which a classification system derives automatically, from MODIS data, cloud and land cover details classified in up to 57 classes.

With the fine spatial resolution, our products are suitable for global- to local-scale air monitoring applications. The 10x10 km^2 resolution of the MODIS standard aerosol product has been increased to 3x3 km^2, providing more details of the emissions sources as well as of the AOT distribution map. Furthermore, the higher spatial resolution and the usage of the classification system have increased the total number of retrieved pixels in the whole scene, especially over land areas and along coastlines.

Moreover, integrated surface information is useful in analysis and improvements of DDV algorithm that uses dark pixels to derive AOT information over land. As shown in the second validation results, the DDV algorithm has performed well on most of dark classes but this approach has limitations on some bright and special dark areas in which improvement is necessary.

In future, we will continue to carry out the validation that compares the results derived by satellite-based and by AErosol RObotic NETwork (AERONET) ground-based measurements in order to verify the consistency of satellite-based AOT products. Besides, we will be aiming at increasing the spatial resolution up to 1x1 km^2 and improving the aerosol retrieval algorithm over bright and dark surfaces.

REFERENCES

Ackerman, S.A., Strabala, K.I., Menzel, W.P., Frey, R.A., Moeller, C.C., Gumley, L.E., 1998. Discriminating clear sky from clouds with MODIS. In: *Journal of Geophysical Research*, 103(D23), pp. 141-157.

Baraldi, A., Puzzolo, V., Blonda, P., Bruzzone L., and Tarantino C., 2006. Automatic Spectral Rule-Based Preliminary Mapping of Calibrated Landsat TM and ETM+ Images. In: *IEEE Transactions on Geoscience and Remote Sensing*, 44(9), pp. 2563-2586.

Dozier, J., 1989. Spectral Signature of Alpine Snow Cover from the Landsat Thematic Mapper. In: *Remote Sensing of Environment*, 28, pp. 9-22.

Environmental Protection Agency, 1999. Air Quality Index Reporting. In: *Federal Register - Rules and Regulations*, 64(149), pp. 42529-42549.

Gupta, P., Christopher, S.A., Wang, J., Gehrig, R., Lee, Y., and Kumar, N., 2006. Satellite remote sensing of particulate matter and air quality assessment over global cities. In: *Atmospheric Environment*, 40, pp. 5880-5892.

Irish, R.R., 2000. Landsat 7 automatic cloud cover assessment (ACCA). In: *SPIE-Algorithms Multispectral, Hyperspectral, and Ultraspectral Imagery VI*, 2049, pp. 348-355.

Kaufman, Y. J. , Tanré, D., 1997. Algorithm for Remote Sensing of Tropospheric Aerosol from MODIS. *MODIS ATBD*.

Kaufman, Y.J., Wald, A.E., Remer, L.A., Bo-Cai, G., Rong-Rong, L., Flynn, L., 1997. The MODIS 2.1-μm channel-correlation with visible reflectance for use in remote sensing of aerosol. In: *IEEE Transactions on Geoscience and Remote Sensing*, 5, pp. 1286-1298

Lary D.J., Remer, L.A., MacNeill, D., Roscoe, B., and Paradise, S., 2009. Machine Learning and Bias Correction of MODIS Aerosol Optical Depth. In: *IEEE Geoscience and Remote Sensing Letters*, 6(4), pp. 694-698.

Li, C., Lau, A.K.H, Mao, J., and Chu, D.A., 2005. Retrieval, Validation, and Application of the 1-km Aerosol Optical Depth From MODIS Measurements over Hong Kong. In: *IEEE Transactions on Geoscience and Remote Sensing*, 43(11), pp. 2650-2658.

MEEO S.r.l., 2009a. PM MAPPER System Description, Issue 1.1. Internal Report, unpublished. If requested, can be delivered upon agreement from the sponsor of the project.

MEEO S.r.l., 2009b. SOIL MAPPER System and Products Description, Issue 4.6, Internal Report, unpublished. If requested, can be delivered upon agreement from the sponsor of the project.

Oo M.M., Hernandez, E., Jerg, M., Moshary, B.M.G., and Ahmed, S.A. 2008. Improved MODIS Aerosol Retrieval using Modified VIS/MIR Surface Albedo Ratio over Urban Scenes. In: *IEEE International Geoscience and Remote Sensing Symposium, 2008*, 3, pp. 977-979.

Remer, L. A, Tanré, D., and Kaufman, Y. J., 2004. Algorithm for remote sensing of tropospheric aerosol from MODIS: Collection 5. *MODIS ATBD*.

Salomonson, V. V., Barnes, W. L. et al., 1989. MODIS: advance facility instrument for studies of the Earth as a system. In: *IEEE Transactions on Geoscience Remote Sensing*, 27(2), pp. 145-153.

Wang, J., and Christopher, S.A., 2003. Intercomparison between satellite-derived aerosol optical thickness and PM2.5 mass: Implications for air quality studies. In: *Geophysical Research Letters*, 30(21), 2095.

CHANGE VISUALIZATION THROUGH A TEXTURE-BASED ANALYSIS APPROACH FOR DISASTER APPLICATIONS

D. Tomowski[a, *], S. Klonus[a], M. Ehlers[a], U. Michel[b], P. Reinartz[c]

[a] Institute for Geoinformatics and Remote Sensing, University of Osnabrueck, 49076 Osnabrueck, Germany - (dtomowski, sklonus, mehlers)@igf.uni-osnabrueck.de
[b] University of Education, Department of Geography, Heidelberg, Germany (e-mail: michel@ph-heidelberg.de)
[c] German Aerospace Center DLR, Remote Sensing Technology Institute, Wessling, Germany - peter.reinartz@dlr.de

KEY WORDS: Change Detection, Comparison, Visualization, Distaster, Texture

ABSTRACT:

A fast detection and visualization of change in crisis areas is an important condition for planning and coordination of help. The availability of new satellites with high temporal resolution (e.g. RapidEye) and/or high spatial resolution (e.g. Quickbird) provides the basis for a better visualization of multitemporal change. For automated change detection, a large number of algorithms has been proposed and developed. This article describes the results of four texture based change detection approaches that were applied to satellite images of the Darfur crisis region. In our methodology we calculate firstly different texture characteristics ("energy", "correlation", "contrast" and "inverse distance moment"), for a whole image at two (or more) different times. The second step is to test the capability of known change detection methods (image-differencing, image-ratioing, regression analysis and principal component analysis) to visualize the change of settlement areas through these texture characteristics and texture images, respectively. The comparison of different texture characteristics with different change detection methods shows that best results can be obtained using a selective bitemporal principal component analysis with the texture feature "energy".

1. INTRODUCTION

A fast detection and visualization of change in crisis areas is an important condition for planning and coordination of help (Kuenzel, 2007). Remote sensing images offer an excellent means for the rapid detection and analysis of change. Consequently, many algorithms for automated change detection have been proposed, developed, and tested. An overview and comparison of different techniques can be found, for example, in Singh (1989), Macleod and Congalton (1998), Mas (1999), Lu et al. (2003), Coppin et al. (2004), or Jianyaa et al. (2008). However, a "best algorithm" for the automated detection of changes for all applications has yet to be developed if this is at all possible (Niemeyer and Nussbaum, 2006). There exist a wide range of different methods with different grades of flexibility, robustness, practicability and significance (Niemeyer and Nussbaum, 2006). These methods can be divided into three categories (Mas, 1999): Image enhancement methods, multitemporal analysis and post classification comparison. Other approaches combine these methods with each other or have a completely novel methodology.

The image enhancement methods are based on unclassified image data, which combine the data mathematically to enhance the image quality (Jensen, 1986). Examples are image-differencing, image-ratioing, principal component analysis (pca), or regression analysis.

Multitemporal analysis (Coppin et al., 2004) is based up on an isochronic analysis of multitemporal image data. This means that n bands of an image at time t_1 and n bands of an image at time t_2 of the same area are merged to form a new image with 2n bands to extract the changed areas in this merged picture (Khorram et al., 1999).

Post classification analysis is based on a comparison of two independent classification results for at least two point of times t_1 und t_2. This method allows the determination of the kind of change from one class to another class.

The high number of published scientific results in the case of combined and novel change detection methods is a clear indication of the importance of this research topic. For example, Prakash and Gupta (1998) combine the image-differencing approach with vegetations indices. Lu et al. (2005) use the image-differencing method together with a principal component analysis. Dai and Khorram (1999) employ neuronal networks whereas Foody (2001) and Nemmour and Chibani (2006) make use of fuzzy-set theory for change detection. Other approaches are based on object-based image analysis (see, for example, Im et al. 2008 or Lohmann et al. 2008).

In our approach, we use image-enhancement methods not for the spectral reflectance values but for different texture characteristics which produce a higher degree of robustness for the change detection and analysis.

2. THEORETICAL BACKGROUND

This section introduces the selected change detection methods for comparison and explains the calculation of the texture features.

2.1 Change detection methods

The first introduced method is image-differencing which is easy to understand and to implement (Singh, 1989). The method is based on calculating the grey value differences. For every pair of grey values at the same location in different points in time the difference is calculated. If the resulting values are unchanged after the algebraic subtraction or do not exceed a

* Corresponding author

pre-determined threshold no change has occurred. The degree of change, on the other hand, is determined by the grey value differences. Very similar is the image-ratio method. For every pair of grey values at the same location in different points in time the ratio of the two values is calculated. Both methods vary through different combinations of spectral bands, the choice of the correct thresholds or different available spectral resolutions. Especially the choice of a correct threshold level is a critical factor (Jensen, 1986), because of a time consuming manual interpretation and the integration of a priori knowledge in the analysis process.

The regression analysis is a statistical method to determine the kind and grade of coherence of features (Mueller, 2000). The idea is to express the relation of a dependent feature to one or more independent variables. In our applied regression method of change detection, pixels from the first time are assumed to be a function of the time t_2 pixel. This relation is expressed through a regression function. The type of regression function is based on the used data or a specific application. In our work we use a least squares method to calculate the regression function. After that, the predicted value (t_2) from the regression line can be subtracted from the value in t_1 to determine the change (for more details see Singh, 1989).

The principal component analysis is a statistical method to calculate new synthetic data space. With this approach, it is possible to intensify wavelength dependent material specific differences. Detailed explanations of this method can be found in Bahrenberg et al. (1992) or Schowengerdt (2007). For change detection, the principal component analysis can be used in different ways (Nussbaum and Menz, 2008). In our study, we employ a selective bitemporal PCA. This means that in a twodimensional feature space two bitemporal spectral bands of the same location are analyzed (Figure 1).

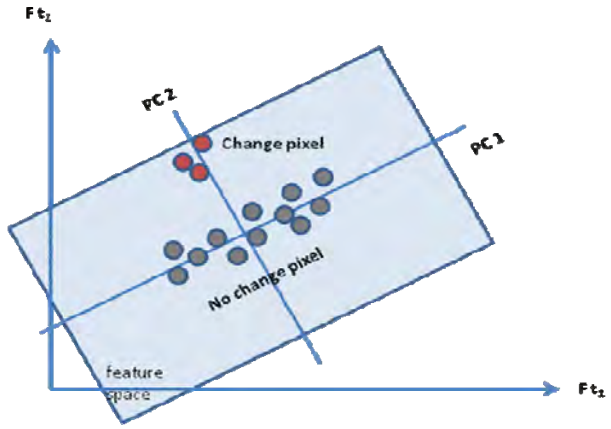


Figure 1. Change detection through bitemporal selective PCA

As result all grey values are located in relation to the principal components. The unchanged pixels have a high correlation with the first principal component in contrast to the change pixels. As a consequence, the first principal component contains the unchanged information und the second component the change information (Macleod and Congalton, 1998).

2.2 Calculation of texture features

The calculation of the texture features is based on the grey-level co-occurrence matrix (GLCM) (Haralick et al., 1973; Haralick and Shapiro, 1992) which represents a second order (Sali and Wolfson, 1992).

The main idea is that settlement areas have higher texture values as non-settlement areas. Many publications of settlement analysis through remote sensing techniques prove these fact (see, for example, Myint, 2007; Steinnocher, 1997; Ehlers and Tomowski, 2008). Here, the GLCM that examine the spectral as well as the spatial distribution of grey values in the image forms the basis for the texture calculation. A GLC matrix describes the likelihood of the transition of the grey value i to the grey value j of two neighbouring pixels (Tomowski et al., 2006). During a calculation of a GLCM, the frequency of all possible grey value combinations of two neighbouring grey values with a defined displacement vector (see figure 2) for both neighbours in both directions of a specific direction (0^0, 45^0, 90^0 or 135^0) can be counted.

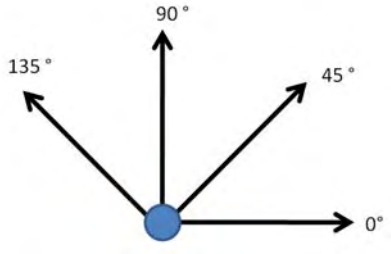


Figure 2. Possible Directions for the displacement vector

This step can be applied for four possible directions. The calculation of the average of these four matrices for every element leads to a direction independent symmetric matrix.

Finally, to calculate the likelihood of a grey value transition, every value in this matrix is divided through the maximum number of all possible grey value transitions (eq. 1):

$$P_{i,j} = \frac{V_{r,c}}{\sum_{r,c=0}^{N-1} V_{r,c}} \qquad (1)$$

where V = value in the symmetric GLCM
r, c = row and column number
N = number of rows or columns

The calculation of the GLCM with high radiometric resolution is a very time consuming step. To reduce this effect, Haralick et al. (1973) suggest different texture features which represent the texture characteristic in a single value. Based on this features new texture images can be calculated over a sliding window technique (for more details see, for example, Lehmann et al., 1997). We use the following texture features in our work:

- *Contrast* (eq. 2) is able to detect the intensity differences between a grey value of a pixel and his neighbourhood.
- *Correlation* (eq. 3) expresses the linear coherence between a grey value pixel in relation to the investigated picture and the texture feature
- *Energy* (eq. 4) calculates the sum of the squared elements in the GLCM and describes the homogeneity of the investigated picture.

- *Inverse distance moment* (IDM) (eq. 5). With the application of the IDM, it is possible to distinguish between heterogeneous and partially homogeneous non-settlement areas.

$$\sum_{i,j=0}^{N-1} P_{i,j}(i-j)^2 \quad (2)$$

$$\sum_{i,j=0}^{N-1} P_{i,j}\left[\frac{(i-\mu_i)(j-\mu_j)}{\sqrt{(\sigma_i^{\,2})(\sigma_j^{\,2})}}\right] \quad (3)$$

$$\sum_{i,j=0}^{N-1} P^2{}_{i,j} \quad (4)$$

$$\sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2} \quad (5)$$

where $\mu_i = \sum_{i,j=0}^{N-1} i(P_{i,j})$

$\mu_j = \sum_{i,j=0}^{N-1} j(P_{i,j})$

$\sigma_i = \sqrt{\sum_{i,j=0}^{N-1} P_{i,j}(i-\mu_i)^2}$

$\sigma_j = \sqrt{\sum_{i,j=0}^{N-1} P_{i,j}(i-\mu_j)^2}$

i , j = grey values

3. IMPLEMENTATION

In this section we introduce our test area. After that, we explain our approach and show the results of our comparison.

3.1 Test area und image data

The study area is located in Sudan. The village Shangil Tobay is located in North Darfur, and was one of seventeen villages in this region, which was attacked and destroyed since 2004 in the Darfur conflict. The conflict in Darfur is a dispute between different ethnic groups, local militia (Janjawid) and the Sudanese government. As is usual the case in these conflicts, the main suffering is sustained by the civilian population. More than 200,000 people have already died; more than 2 Million people have been displaced. The region around Shangil, a sparsely populated area, is the home of several thousands of displaced civilians. The attacks on this village took place between 2005 and 2006. To document the dimension of the inhumanities, Amnesty International (2009) and AAAS (2009) maintain web sites that shows satellite images of the affected regions before and after an attack on several villages. With the permission of the satellite company Digital Globe, we were able to use preprocessed georeferenced Quickbird data before and after an attack for our change analysis. Figures 2 and 3 show the study area on 10 March 2003 and 18 December 2006, respectively. For our tests, we used a subset of 512 pixel * 512 pixel with a 0,6 m ground pixel size and a radiometric resolution of 8 bit. A visual comparison and overlay of the existing man-made structures shows a high congruence between both images, so that a new co-georeferencing was not necessary and the problem of possible pseudo change was negligible.
The selected subset is a village in the east of Shangil Tobay. It is visible, that the nearly the whole village was demolished; only a few buildings are not destroyed. Some of the structures were completely wiped out; some were burnt down and are still partly visible.

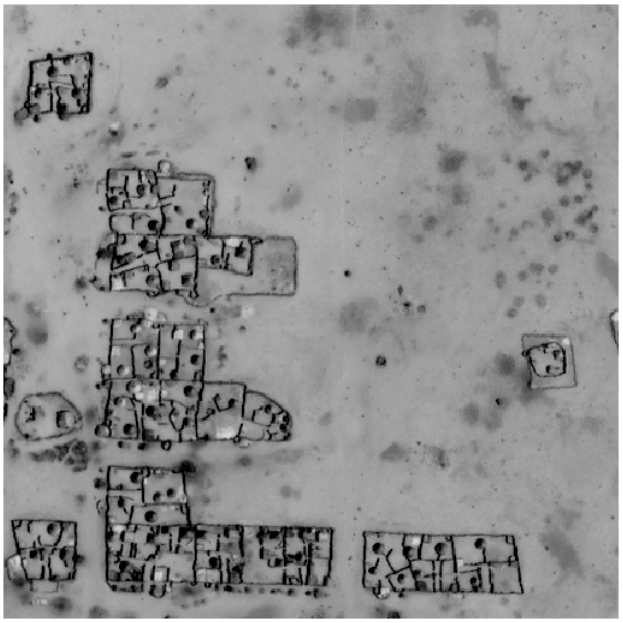


Figure 3. Subset of the Shangil village before destruction (©Digital Globe 2003)

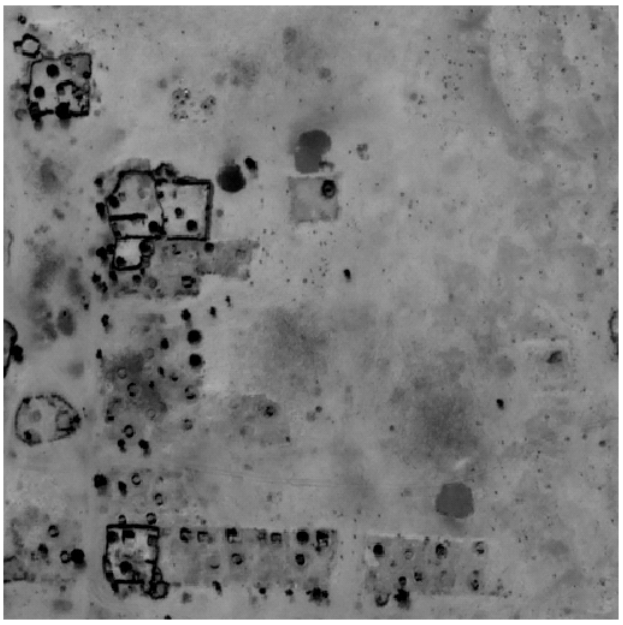


Figure 4. Subset of the Shangil village after destruction (©Digital Globe 2006)

3.2 Change Detection Concept and Design

Figure 5 shows the general steps that are the prerequisite for the change detection analysis. To apply our change detection methods, firstly all 3 bands (RGB) from the images taken at t_1 and t_2 were averaged to obtain a single band. The different texture features are calculated from the averaged data and then submitted to the different change detection analysis methods (image-differencing, image-ratioing, regression analysis and PCA). Basis for the calculation of the texture features are the images at the specified time t_1 und t_2. After the texture calculation, we derived the GLCM (8 bit) for every image. Based on the GLCM we calculated the texture features *contrast*, *correlation*, *IDM* and *energy* for every point in time with differently sized windows (ranging from 3 x 3, to 13 x 13 windows). Using these texture features, we then performed the four change detection methods explained above.
Because of the different illumination conditions in the two images (see figure 2 and 3) we also included in a second processing step a histogram matching to both image data sets to improve the contrast and to harmonize the grey value distribution of the images (Yang and Lo, 2000).

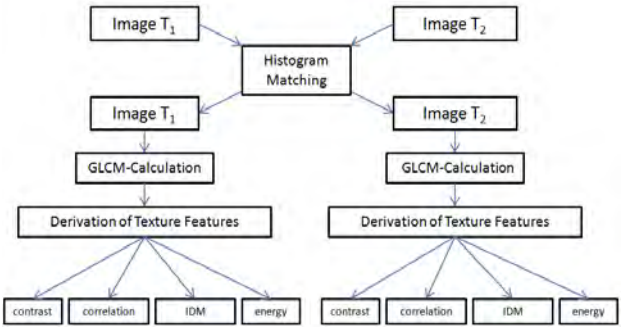


Figure 5. Steps for the derivation of texture image data for change detection

Best results were obtained using a 13 x 13 window for the calculation of the texture features after the histogram matching. Again, we performed the change detection analysis four times (image-differencing, image-ratioing, regression analysis and PCA) with all derived texture images.

3.3 Results and interpretation

As result, we generated 16 change images (with 4 kinds of texture features and 4 change detection methods). In this paper, we will concentrate on best result and present only a short summary for the other findings. As our goal was the visualization of changes, we did not perform a quantitative accuracy analysis but concentrated on a qualitative visual interpretation (see table 8).
The image-differencing method showed the worst results for all four texture images. Therefore we will not pursue this methods in further studies.
The texture based image-ratio approach did not produce satisfying results in the case of the texture features *contrast* and *correlation*. Certainly, with the IDM and the energy feature a partial detection of change in the settlement areas is visible.
The texture based regression analysis led to partly suitable results; problems were the imprecise location of building contours. It may have some usefulness, however, in object-based image analysis.

The results for the texture based PCA with the images derived from the features *contrast*, *correlation* und *IDM* were not satisfying. However, bitemporal PCA for the texture feature *energy* proved to be the best results as far as visual appearance is concerned.

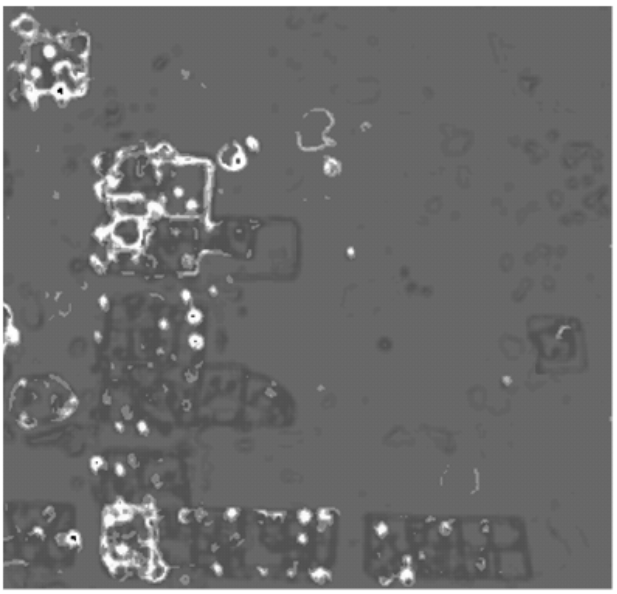


Figure 6. Results of texture based PCA with the texture feature *energy*

It is possible to visualize changed (dark grey) unchanged settlement areas (bright grey) and unchanged non-settlement-areas (medium grey) without determine thresholds (see figure 6):

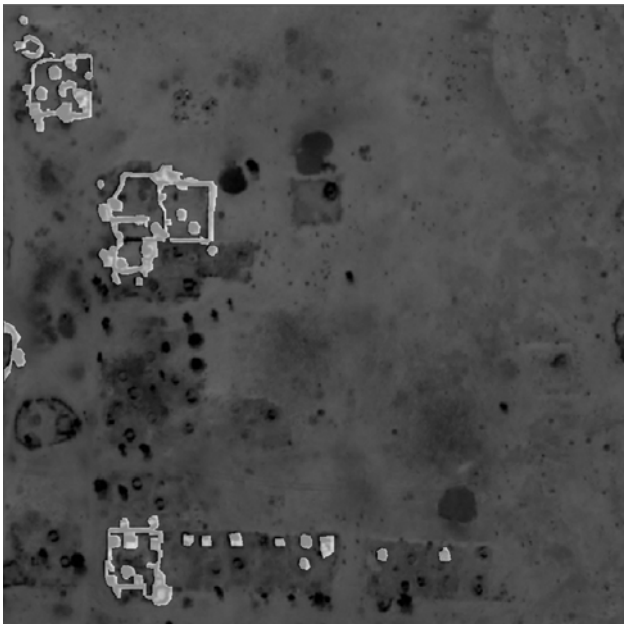


Figure 7. Manual digitized unchanged man-made structures and the subset of the Shangil village after destruction
(©Digital Globe 2006)

The comparison with a manual digitized classification (based on figure 4) of the unchanged man-made structures (buildings, fences) shows (see figure 7), that the algorithm visualizes

nearly all unchanged buildings, but also some small areas with high texture values, that have changed (ruins and stone formations). However, it is easy to distinguish between destroyed and not destroyed structures.
Table 8 shows the overall results of the qualitative assessment respectively visual interpretation for the different methods and texture features. The table compares and assesses the change results of the different change detection methods (left column) in relation to the four texture images (upper row).

	Con-trast	**Correlation**	**IDM**	**Energy**
Image-differencing	-	-	-	-
Image-ratio	-	-	0	0
Regression	0	0	0	0
PCA	-	-	-	+

Table 8. Assessment of the texture based change detection result: Evaluation: „+“ = good visualization, „0“ = partly visualization and „-“ = bad visualization of changes

4. CONCLUSIONS

A comparison of four different texture characteristics *contrast*, *correlation*, *inverse distance moment* (IDM) and *energy* that were used for four different change detection methods (image-differencing, image-ratioing, regression and PCA) shows that a combination of a bitemporal principal component analysis with the texture feature *energy* displays the best results for the visualization of change. Image of a village in the Darfur region before and after destruction were used for our study. However, the used test area was ideal to show the possibilities of texture-change analysis, because most of the buildings were completely destroyed and no texture was present on the locations of the destroyed buildings.
With the texture-based change detection approach, it is possible to visualize changed and unchanged settlement areas without the determination of thresholds. In our future work we will test this approach with other satellite images and airborne data and to show the transferability of this method.

REFERENCES

AAAS, 2009. http://shr.aaas.org/geotech/darfur/darfur.shtml, Homepage of AAAS, accessed: October 2009.

Amnesty International, 2009. http://www.eyesondarfur.org, Homepage of Amnesty International, accessed: October 2009.

Bahrenberg, G., Giese, E. and Nipper, J., 1992. *Statistische Methoden in der Geographie: Multivariate Statistik* (Vol. 2). Vieweg+Teubner, Wiesbaden.

Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B. and Lambin, E., 2004. Review Article. Digital change detection methods in ecosystem monitoring a review. *International journal of remote sensing, 25*(9), pp. 1565–1596.

Dai, X. and Khorram, S. ,1999. Remotely Sensed Change Detection Based on Artificial Neural Networks. *Photogrammetric engineering and remote sensing, 65*(10), pp. 1187–1194.

Ehlers, M. and D. Tomowski, 2008. On Segment Based Image Fusion, in: Blaschke, T., S. Lang and G. Hayes (Eds.), Object-Based Image Analysis – Spatial Concepts for Know-ledge-Driven Remote Sensing Applications, Springer Lecture Notes in Geoinformation and Cartography, pp. 735-754.

Foody, G. M., 2001. Monitoring the Magnitude of Land-Cover Change around the Southern Limits of the Sahara. *Photogrammetric Engineering and Remote Sensing, 67*(7), pp. 841–847.

Haralick, R. M., Shanmugam, K. and Dinstein, I., 1973. Textural features for image classification. *IEEE Trans. Syst., Man, Cybern.*, (SMC-3), pp. 610–621.

Haralick, R. M. and Shapiro, L. G., 1992. *Computer and Robot Vision - Volume 1*. Addison-Wesley, Reading.

Im, J., Jensen, J. R. and Tullis, J. A., 2008. Object-based change detection using correlation image analysis and image segmentation. *International Journal of Remote Sensing, 29*(1-2), pp. 399–423.

Jensen, J. R., 1986. *Introductory Digital Image Processing: A Remote Sensing Perspective.* Prentice-Hall, Englewood Cliffs, New Jersey.

Jianyaa, G., Haiganga, S., Guoruia, M. and Qimingb, Z., 2008. A Review of multi-temporal remote sensing data change detection algorithms. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Beijing, China, Vol. XXXVII, Part B7, pp. 757-762.

Khorram, S., Biging, G. S., Chrisman, N. R., Colby, D. R., Congalton, R. G. and Dobson, J. E. et al., 1999. *Accuracy Assessment of Remote Sensing Derived Change Detection.* American Society for Photogrammetry and Remote Sensing, Bethesda.

Kuenzel, W., 2007. Kartographie für den schnelleren Rettungseinsatz bei Naturkatastrophen: GIS bei den Vereinten Nationen und das Erdbeben in Pakistan. In E. Theile (Eds.), *Kartographische Schriften: Vol. 14. Kartographie - Ihnen werden wir's zeigen. Symposium 2007, 14 - 16. Mai 2007 Koenigslutter am Elm.* Kirschbaum, Bonn.

Lehmann, T., Oberschelp, W., Pelikan, E. and Repges, R., 1997. *Bildverarbeitung für die Medizin: Grundlagen, Modelle, Methoden, Anwendungen:* Springer, Berlin.

Lohmann, P., Hoffmann, P. and Müller, S., 2008. Updating GIS by object-based change detection. In J. Schiewe and M. Ehlers (Eds.), *gi-reports@igf - Geoinformatics paves the Highway to Digital Earth: On the occasion of the 60th birthday of Professor Manfred Ehlers* (Vol. 8, 1st ed., pp. 81–86). Osnabrueck.

Lu, D., Mausel, P., Batistella, M. and Moran, E., 2005. Land-cover binary change detection methods for use in the moist tropical region of the Amazon: a comparative study. *International Journal of Remote Sensing, 26*(1), pp. 101–114.

Lu, D., Mausel P., Brondízio E. and Moran E., 2003. Change detection techniques. *International Journal of Remote Sensing, 25*(12), pp. 2365–2407.

Macleod, R. D. and Congalton, R. G.,1998. A Quantitative Comparison of Change-Detection Algorithms for Monitoring Eelgrass from Remotely Sensed Data. *Photogrammetric Engineering and Remote Sensing, 64*(3), pp. 207–216.

Mas, J. –F, 1999. Monitoring land-cover changes: a comparison of change detection techniques. *International Journal of Remote Sensing, 20*(1), pp. 139–152.

Mueller, B., 2000. Statistische Auswertungen von Untersuchungsergebnissen. In H. Barsch, K. Billwitz, and H.-R. Bork (Eds.), *Perthes Geographie Kolleg:Arbeitsmethoden in Physiogeographie und Geooekologie.* Gotha: Klett-Perthes, pp.407–449.

Myint, S.W. , 2007. Urban Mapping with Geospatial Algorithms. In Q. Weng and D. A. Quattrochi (Eds.), *Urban Remote Sensing*. CRC Press, Boca Raton, pp.109–136.

Nemmour, H. and Chibani, Y., 2006. Fuzzy neural network architecture for change detection in remotely sensed imagery. *International Journal of Remote Sensing, 27*(3-4), 705–717.

Niemeyer, I. and Nussbaum, S., 2006. Automatisierte Detektion, Klassifizierung und Visualisierung von Veränderungen auf der Grundlage von Fernerkundungsdaten. In I. Niemeyer, A. Sroka and R. Wittenburg (Eds.), *Schriftenreihe des Institutes für Markscheidewesen und Geodäsie an der Technischen Universität Bergakademie Freiberg: Tagungsband des 7. Geokinematischen Tages vom 11. und 12. Mai 2006 in Freiberg.* Glueckauf, Essen, pp. 248–257.

Nussbaum, S. and Menz, G., 2008. *Object-Based Image Analysis and Treaty Verification: New Approaches in Remote Sensing - Applied to Nuclear Facilities in Iran.*Springer Science+Business Media B.V., Dordrecht.

Prakash, A. and Gupta, R. P., 1998. Land-use mapping and change detection in a coal mining area - a case study in the Jharia coalfield, India. *International Journal of Remote Sensing, 19*(3), pp. 391–410.

Sali, E. and Wolfson, H., 1992. Texture classification in aerial photographs and Satellite data. *International Journal of Remote Sensing, 13*(18), pp. 3395–3408.

Schowengerdt, R. A., 2007. *Remote Sensing: Models and methods for image processing* (3. ed.). Elsevier Academic Press, Amsterdam.

Singh, A., 1989. Digital change detection techniques using remote-sensed data. *International Journal of Remote Sensing, 10*(10), pp. 989–1003.

Steinnocher, K. T., 1997. Texturanalyse zur Detektion von Siedlungsgebieten in hochaufloesenden panchromatischen Satellitenbilddaten. In *AGIT IX, 2.-4. Juli 1997, Salzburger Geographische Materialien.* Vol. 26, pp. 143–152.

Tomowski, D., Ehlers, M., Michel, U. and Bohmann, G., 2006. *Objektorientierte Klassifikation von Siedlungsflaechen durch multisensorale Fernerkundungsdaten*, gi-reports@igf, Vol. 3, University of Osnabrueck, Osnabrueck.

Tomowski, D., Michel, U., Ehlers, M. and Bohmann, G., 2006. Siedlungsflaechenklassifikation mittels objektorientierter Datenfusion durch multisensorale Fernerkundungsdaten. In U. Michel and K. Behncke (Eds.), *gi-reports@igf: Vol. 5. Veroeffentlichungen des AK Fernerkundung 2006.* University of Osnabrueck, Osnabrueck, pp.123–135.

Yang, X. and Lo, C. P., 2000. Relative Radiometric Normalization Performance for Change Detection from Multi-Date Satellite Images. *Photogrammetric Engineering & Remote Sensing 66*(8), 967–980.

ACKNOWLEDGEMENTS

The presented work was done in the research project „Automated Change Detection methods for disaster applications“ at the Institute for Geoinformatics and Remote Sensing (IGF) at the University of Osnabrueck.

REMOTE SENSING, GEOGRAPHIC INFORMATION SYSTEMS AND SHANNON'S ENTROPY: MEASURING URBAN SPRAWL IN A MOUNTAINOUS ENVIRONMENT

L. C. O. Verzosa [a, *], R. M. Gonzalez [a]

[a] Department of Geodetic Engineering, University of the Philippines, Diliman, Quezon City 1101, Philippine
lc_verzosa@yahoo.com, rmgonzalez@up.edu.ph

KEY WORDS: Geographic Information Systems, Photogrammetry, Urban, Measurements, Image

ABSTRACT:

Urban sprawl, or the unplanned and uncontrolled spreading out of built-up areas, causes problems in the allocation of basic needs and increases risk to life and property in the face of disasters. The integration of remote sensing and geographic information systems is used in adopting Shannon's entropy to measure urban sprawl. Shannon's entropy is an index used here in quantifying the degree of dispersion or concentration of built-up areas. This study in the mountainous city of Baguio in northern Philippines shows that together with remote sensing, geographic information systems and photogrammetric techniques, built-up concentration can be identified and quantified from time series of aerial photographs and satellite images; this facility can assist in monitoring the growth of built-up areas and in drafting measures and policies to address urban sprawl's imminent effects.

1. INTRODUCTION

1.1 The Problem with Urban Sprawl

Urbanization is needed for development (Bekele, 2005). However, growth and spread of built-up areas that come with urbanization can become uncontrolled and irregular over time, such that isolated tracts of land are encroached upon. This is known as "urban sprawl" (Yeh and Li, 2001 p.83; Mujtaba, 1994 p.2) and such is detrimental to the efficient functioning of developed areas. Cities become congested due to overpopulation, and resources become limited in catering to the different needs of the people. Moreover, urban sprawl presents greater risk and damage to life and property in the occurrence of calamities and disasters. It endangers the living condition of inhabitants and puts ecosystems in jeopardy by compromising biodiversity (Yeh and Li, 2001). Thus, urban sprawl studies are done to help monitor the spread of built-up areas and quantify the sprawl in order to determine the trend, the extent, and avert the associated complications (Yeh and Li, 2001; Sudhira *et al.*, 2004; Jat *et al.*, 2007).

1.2 Urban Sprawl Studies

One of the measures commonly used due to its robustness in urban sprawl measurement is Shannon's entropy (Yeh and Li, 2001); it is an index that determines the distribution of built-up as a function of the area of built-up within a defined spatial unit (Jat *et al.*, 2007). It characterizes the pattern – dispersed or concentrated – of built-up over time that can help officials to identify which area is being used inefficiently (Yeh and Li, 2001). Moreover, entropy values can be factored into the analysis of risk as a component of vulnerability in the risk equation: *risk = hazard x vulnerability* (Castellanos-Abella, 2008). Entropy gives the distribution of the loss to the built-up, which signifies population and property.

Previous studies that use Shannon's entropy take into account the horizontal spread of built-up on a given area (Yeh and Li, 2001; Jat *et al.*, 2007). However, as time progresses, people begin to develop residences vertically, constructing structures of several storeys high in order to maximize available habitable space especially in mountainous regions (Reddy, 1996). This makes the risk per unit of space more acute. Thus, both horizontal distribution and vertical growth need to be measured to determine the full scope of the implications of urban sprawl in the affected area.

Entropy calculation is based on area computation; this is best facilitated by the integration of spatial measurement facility correspondingly offered by remote sensing (RS), geographic information system (GIS) and photogrammetric techniques (Sudhira *et al.*, 2004; Ayhan *et al.*, 2008). Satellite images and aerial photographs provide data about the physical state of an area at a given time. Remote sensing facilitates image enhancement and classification; photogrammetric techniques, particularly stereo-modelling, enable the determination of heights and elevations. On the other hand, GIS enables proper storage, retrieval, and display of spatial data, as well as provide spatial analysis functions and area computation. The combination of these tools provides the essential data in computing for Shannon's entropy in growth areas; thereby check on emerging trends that may require immediate attention.

1.3 The Study Area

In this study, we apply the principle of Shannon's entropy to measure urban sprawl phenomenon in the highly urbanized mountain city of Baguio, located in the Province of Benguet in the northern Philippines (see Figure 1).

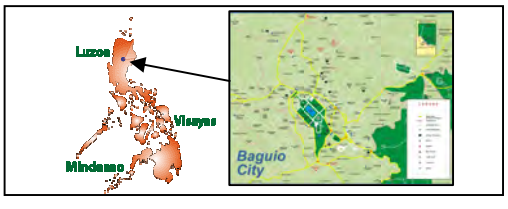


Figure 1. Location Map of Baguio City
(Philtravel Center; CLUP)

* Corresponding author

Baguio is a plateau at 1400m altitude whose limited land area was originally designed for 30,000 people (MTDP 2005-2010); but the 2007 census shows that it is now actually housing some 301,926 inhabitants (NSO 2007). Baguio City is the Philippine's "Summer Capital" as its temperature is generally 8-10°C lower than those of the lowlands. Baguio City's cool climate attracts settlers such that whole mountainsides have developed into housing projects. With its limited space, high-rise structures are dominating Baguio City's landscape. This poses a big risk to its inhabitants due to its steep terrain, for being traversed by the northern splay of the Philippine Fault Zone, and for being in a wet region; Baguio City holds the world record for a 24-hour rainfall at 1,200mm (Jennings, 1950). With such confluence of environmental characteristics, Baguio City is highly prone to landslides (Orense, 2003).

Figure 2. Urban Sprawl along the Slopes of Baguio City (LCVerzosa, 2009)

To capture both the horizontal and vertical configurations of urban sprawl in Baguio City's mountainous landscape, this paper reports about the integration of remote sensing, GIS and photogrammetric techniques in conjunction with Shannon's Entropy to quantify the developed areas over time. We hope that this combination of technologies and procedures can realistically illustrate the spread of development in Baguio City and help its local government to effectively monitor and plan to mitigate the possible effects and implications of urban sprawl to its community.

2. METHODOLOGY

2.1 Shannon's Entropy: *Horizontal* and *Vertical* Entropy

Shannon's entropy originated from information theory as a measure of uncertainty of conveyed information over a noisy channel (Jat *et al.*, 2007; Bailey, 2009). The larger the value of Shannon's entropy, the higher is the uncertainty of information conveyed (Bailey, 2009). High entropy is the most probable yet least predictable state that leads to disorder. The same is true for urban sprawl. Thus, Shannon's entropy is convenient in measuring urban sprawl (Yeh and Li, 2001).

In urban sprawl, Shannon's entropy is the measure of the degree of dispersion or concentration of a random geographical variable, i.e. the built-up. It is given by (Yeh and Li, 2001):

$$H_n = \sum_i^n p_i \log(\tfrac{1}{p_i}) \qquad (1)$$

where p_i is the probability or proportion of occurrence of a phenomenon in the i^{th} spatial unit out of n units, and thus, is given by:

$$p_i = x_i / \sum_i^n x_i \qquad (2)$$

where X_i is the area of built-up at the i^{th} unit.

Entropy values range from a minimum of 0 – representing a concentrated pattern, to a maximum of *log n* – denoting a dispersed distribution. The value of *log n* is significant as it can be treated as a critical limit or threshold to the expansion of an area. *Log n* can be computed for any region; and every year, the entropy value can be monitored and compared to the previous in order to avoid reaching the threshold. Thus, planners and local officials would be able to have a scientific and quantifiable description of the urban situation for drafting policies that take into account threshold value.

When entropy values for succeeding years are investigated, the progression of a random geographic variable – the sprawl of built-up can be monitored. Increasing entropy values would indicate continuous dispersion with built-up highly occurring. Decreasing values would signify that an area is becoming less fragmented and homogenously covered, thus, further occurrence of built-up is less likely to happen.

Equations (1) and (2) above are used to calculate both *horizontal* and *vertical* entropies. The difference lies in determining the area of built-up, x_i, to be considered. *Horizontal* entropy describes the horizontal spread of the built-up area; thus, it accounts for the plane area of the built-up. On the other hand, *vertical* entropy evaluates the vertical development or the rise of building structures; thus, the aggregated floor area of the built-up is computed. In this case, x_i is given by:

$$x_i = \textit{(number of floors) x (area of the building footprint)} \qquad (3)$$

where the number of floors is a function of the building height and the standard floor height.

Table 3 shows the three possible levels at which the Entropies can be computed in this study.

LEVEL	SPATIAL UNIT
City	*Barangay* (village)
Barangay (village)	*Purok* (neighbourhood)
Purok (neighbourhood)	Building

Table 3. Levels of Entropy Computation

For this study, *horizontal* entropy is computed for: (1) the entire city of Baguio using a series of available Landsat images from 1979, 1989, 1992, and 2002 in order to investigate and illustrate the trend and extent of horizontal development with population increase; (2) the *barangay* (village) level, and, (3) the *purok* (neighbourhood) level of Barangay Cresencia Village, one of the severely affected *barangays* during a recent typhoon, Typhoon Parma of October 2, 2009.

Computation of *vertical* entropy at *purok* level of Barangay Cresencia Village is also done in order to investigate the increase in high-rise buildings and analyze its relationship with the *horizontal* spread. Aerial photographs from 2003 are used to determine building heights.

2.2 Data and Software Used

Table 4 shows the available data which are utilized for this study. The Landsat images provide a time series analysis on the spread of built-up areas, while the aerial photographs are interpreted to yield the extent and to determine building heights in the developed areas. Identification of these built-up areas is essential in locating those where the population is concentrated. Kawamura (1971) similarly showed how changes in urban growth are monitored using a series of aerial photographs.

DATA	SOURCE
Landsat Images (1979, 1989, 1992, 2002)	USGS GLOVIS
Aerial Photographs (2003)	F.F. Cruz and Co.
Barangay Boundary Map	Reclaim II Project
Topographic Map	National Mapping and Resource Information Agency

Table 4. The Gathered Data

Satellite image enhancement and classification are done with ENVI 4.3®, while ERDAS-Imagine® is used for stereo-modeling of the aerial photographs. Spatial analysis and area computation are done using GIS software ArcGIS 9.3®.

2.3 Flowchart of the Methodology

Figure 5 illustrates the methodology we developed that combines photogrammetric and digital image processing techniques to delineate and estimate the extent of urban growth areas. GIS is also used in visualizing the spatial spread and distribution of developed areas. In turn, the quantified built-up areas from the aerial photographs and satellite images are used to compute Shannon's entropy.

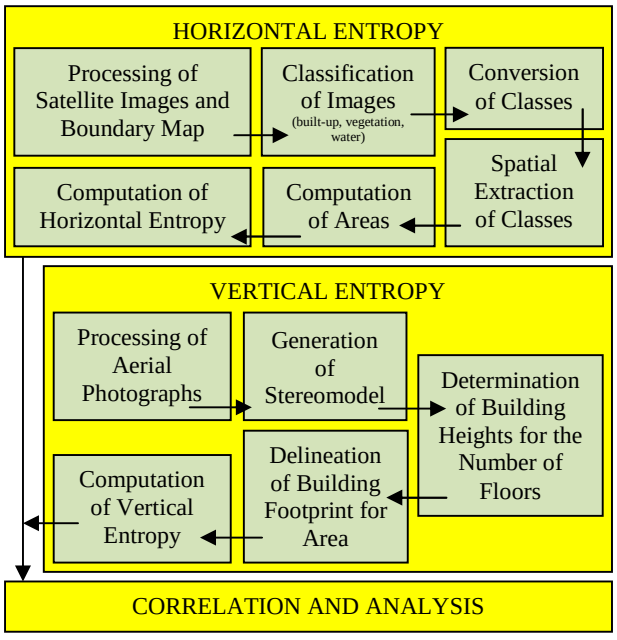


Figure 5. Methodology in the Measurement of Urban Sprawl in a Mountainous Environment

3. RESULTS AND DISCUSSION

3.1 Horizontal Entropy: Baguio City

From the temporal *horizontal* entropy values, all of which are near log N, and the corresponding *barangay*-built-up images of each year (see Fig. 6), it can be seen that the pattern of built-up area manifest a dispersed distribution from the city center (Barangay Kisad-Legarda). The high values of HE also indicate that the city is approaching log N (7.01), which is the critical level of expansion of built-up. However, from 1992 to 2002, there is a significant drop in entropy (from 6.04 – 5.79); this indicates that due to the rapid growth, the area has become homogenously covered by built-up, the space has become very limited and built-up areas have become concentrated within the city limits. However, in Figure 7, a population census from 1995-2000 (NSO, 2007), shows yet an increase from 226,883 to 252,386 (10% growth). This yields a higher population density that, without a corresponding built-up spread, is a sign of vertical development; the people are building high-rise residences. Such a trend necessitates the investigation of *vertical* entropy.

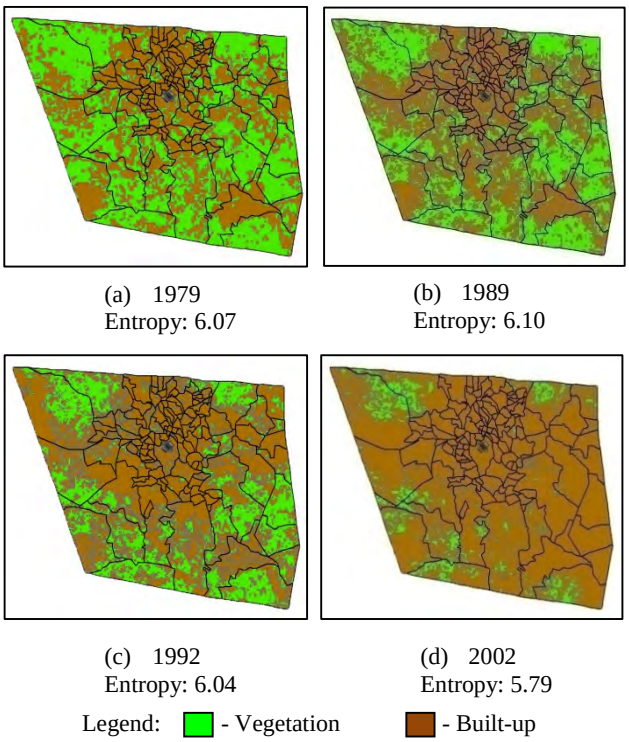


Figure 6. Series of Satellite Imagery and Corresponding Entropy Depicting the Spread of Urban Growth in Baguio City from 1979 to 2002

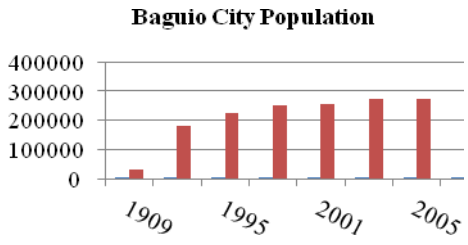


Figure 7. Baguio City Population from 1909-2007 (NSO, 2007)

3.2 Horizontal and Vertical Entropy: Barangay Cresencia Village

From the *horizontal* entropy value (HE) of the whole Cresencia Village in Table 8, it can be seen that this *barangay* has high entropy indicative of sprawl and availability of space where built-up areas may further establish. In the same way, the high *vertical* entropy value (VE) of this *barangay* signifies that built-up, in reality, is more than what is being shown in the horizontal dimension. It reflects a larger dwelling area, as well as population concentration. Thus, the two values would indicate the magnitude of the risk in the event of a disaster or calamity. This analysis is further established in the *purok* level (see Table 9), whose entropy values are near the threshold value.

With the computed *purok* entropy values, an entropy map can be generated using an entropy matrix for a given hazard. As an example, consider the entropy matrix for landslide risk in Figure 10. Entropy values are classified as high or low depending on their nearness to log N (Jat et al., 2007). An entropy value is high when it is nearer to log N, and vice versa. From this matrix, the HE and VE of the *puroks* of Brgy. Cresencia Village are classified and the resulting entropy map is generated in Figure 11. Entropy maps are useful in visualizing the risk corresponding to the horizontal distribution and vertical growth of built-up.

	HE	VE	N (*puroks*)	LOG (N)
Cresencia Village	1.73	1.68	4	2

Table 8. *Horizontal* and *Vertical* Entropies of Barangay Cresencia Village

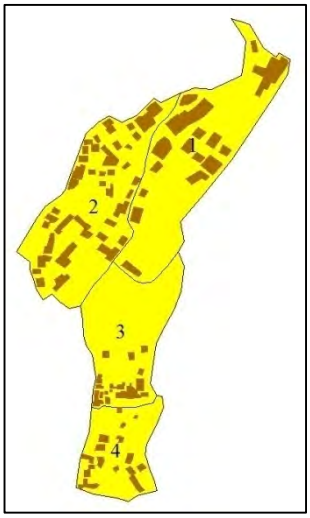
PUROK	HE	VE	N (building)	LOG (N)
1	4.24	3.96	22	4.46
2	5.64	5.43	58	5.86
3	4.05	4.00	21	4.39
4	3.93	3.76	18	4.09

Table 9. *Horizontal* and *Vertical* Entropies of the *Puroks* in Barangay Cresencia Village

VE	*High*	**HIGH RISK** (Concentrated, High Rise)	**RISKY** (Dispersed, High Rise)
	Low	**RISKY** (Concentrated, Low Rise)	**LOW RISK** (Dispersed, Low Rise)
		Low	*High*
		HE	

Figure 10. Entropy Matrix for Landslide Risk

The entropy map in Figure 11 shows that for a landslide risk, the puroks of Brgy. Cresencia Village are classified as risky areas composed of dispersed high-rise buildings.



Legend: - High HE, High VE (dispersed, high rise)
- Built-up

Figure 11. Entropy Map of Barangay Cresencia Village

3.3 The Aftermath of Typhoon Parma

Baguio City experienced severe casualties when the tropical storm Parma hit the city in October 2, 2009. Major landslides and high flooding were experienced in conjunction with the heavy rainfall that reached an average of 927.8 mm on that day (Gonzales, 2009). At the height of the typhoon three major roads going to and from Baguio were closed because they were impassable due to landslides in the area (Mananghaya, 2009). Figure 12 shows snapshots of some of the affected barangays.

Figure 12. Top: Flooding at the City Camp
(L – Noel Godinez, R – Brenda Dacpano; www.nordis.net);
Bottom: Landslide at Cresencia Village
(L – www.daylife.com, R – Danny Durante)

Tables 13 and 14 show the effects of the Typhoon Parma from October 2 – 13, 2009. The reported incidents occurred in different areas of Baguio City such as Brgy. Cresencia Village, Kennon Road and Bokawkan Road.

REPORTED INCIDENTS	NUMBER
Eroded Riprap	25
Soil Erosion	39
Flood	41
Landslide	58
Vehicular Accidents	1
TOTAL INCIDENTS	*164*

Table 13. Summary of Incidents (Pepeng Report, 2009)

Id	BARANGAY/ROADS	Evacuated Inhabitants	Missing	Deaths	Injured
1	Irisan	196	1	15	10
2	Pinsao_Proper	50		2	
3	Pinsao_Pilot	50			
4	Fairview	80			
5	San_Luis_Village	7			
6	Dominican_Hill_Mirador	12			
7	Sto._Rosario	47			
8	Quirino_Magsaysay_Upper	22			
9	Rock_Quarry_Upper			2	
10	Rock_Quarry_Middle			2	
11	Quirino_Magsaysay_Lower	557			
12	City_Camp_Proper	125			
13	Rock_Quarry_Lower	152			
14	City_Camp_Central	256			
15	Queen_Of_Peace	144			1
16	Camp_Allen	90			
17	Cresencia_Village			23	5
18	Gen._Luna_Upper	20			
19	Kias	26			
20	Atok_Trail				2
21	Loakan_Apugan			1	
22	Mines_View	6			
23	Brookspoint			1	
24	Bayan_Park_East				1
25	Aurora_Hill_South_Central				1
26	Bayan_Park_Village				1
27	Aurora_Hill_North_Central				1
28	Bayan_Park_West				1
29	Aurora_Hill_Proper				1
30	Lourdes_Subd_Extension	100			
31	Marcos Highway			4	1
32	Kennon Road				2
33	Kitma			8	

Table 14. Summary of Casualties (Pepeng Report, 2009)

Table 14 shows that Barangay Cresencia Village registered the most number of casualties during the recent typhoon. This reflects the high entropy values in Table 8, which was computed based on a set of 2003 aerial photographs. Therefore, these 2003 estimates, and whenever data is available to compute for the entropy of the succeeding years, would have served as signs that point to the high risk *barangays* in the city; the local government can then focus on high risk *puroks* in each *barangay*. In essence, a more strategic disaster preparedness plan could have been drafted.

The growth of built-up in Baguio City has become uncontrolled over the years, even with its policies that aimed at preventing congestion (CLUP, 2002-2008). The consecutive entropy values over a given period would suggest both the horizontal and vertical patterns of growth of the built-up; thus, better estimations of further growth could be seen and properly addressed in future policies.

In the light of the recent typhoon casualties and damages, the city council of Baguio City saw the need to draft resolutions and ordinances that would answer the city's lack of proper housing policies, disaster preparedness plans and rehabilitation measures. An ordinance is now being proposed for the creation of the City Housing and Resettlement Office to facilitate the "proper formulation, implementation and monitoring of housing and resettlement programs and projects in the city" (Refuerzo, 2009a). This study on the integration of remote sensing and geographic information systems in adopting Shannon's entropy to measure urban could help in giving a quantitative and scientific approach to site selection for housing and relocation projects.

The flooding that occurred at the City Camp Lagoon has been attributed to trash that clogged the sinkholes in the underground water channel (Refuerzo, 2009b). In this case, entropy values, which also reflect population concentration in the area, could have served as estimates of the number of people that will be affected during a flooding incident. The values can also suggest the amount of solid waste that are produced; and which potentially clog the drains when typhoons occur.

The death toll and infrastructure damage brought by Typhoon Parma prompted the Baguio city government to focus on areas found to be geologically hazardous, and to create a wholistic disaster management plan (Refuerzo, 2009c). The *horizontal* and *vertical* estimates of Shannon's entropy reflect the built-up area and population situation of the city; thereby, providing a comprehensive approach to prevent further casualties in the face of natural calamities.

4. CONCLUSION

Integration of photogrammetric, remote sensing, and GIS techniques facilitates delineation, tracking down and monitoring of urban development. RS provides pattern recognition techniques to classify land cover based on their spectral characteristics on satellite images. GIS enables the proper handling of databases necessary for the integration of data from different sources. Photogrammetric techniques for measuring building heights in estimating the vertical population concentration is also sufficient in providing the necessary data for computing Shannon's entropy equation. Entropy values are obtained and this demonstrates how it can be implemented within a GIS to facilitate the measurement and visualize the extent of urban sprawl.

This study shows that entropy is a good indicator in identifying and monitoring land development—that is, dispersion and concentration of built-up areas. Compactness of development (indicated by low entropy) is a sign of vertically built development in the light of continuing increase in population. This imply that high population concentration is being exposed to Baguio City's geologic hazards such as earthquakes and landslides, and more recently, flooding hazards. Moreover, the high values of Shannon's entropy indicate that the city is precariously approaching its critical level. Such a situation is

important to keep track by the city government; in this way, safe urban development can be planned.

REFERENCES

Ayhan, E., Erden O., Gormus, E.T., 2008. Three dimensional monitoring of urban development by means of ortho-rectified aerial photographs and high-resolution satellite images. *Environ. Monit. Assess.*, Vol. 147, Nos. 1-3, pp. 413-421. http://www.springerlink.com/index/A7U5443530663M04.pdf (accessed 20 July 2009).

Bailey, K.D., 2009. Entropy systems theory. *Systems Science and Cybernetics*, Eolss Publishers, Oxford, UK, pp. 152-169. http://www.eolss.net/ebooks/Sample%20Chapters/C02/E6-46-01-04.pdf (accessed 1 December 2008).

Bekele, H., 2005. Urbanization and urban sprawl. http://www.kth.se/polopoly_fs/1.21420!294.pdf (accessed 2 December 2008).

Castellanos-Abella, E.A., 2008. Multi-scale Landslide Risk Assessment in Cuba. *ITC Library Academic Output 2008*. http://www.itc.nl/library/papers_2008/phd/castellanos.pdf (accessed 1 March 2010).

CLUP. 2002-2008. Baguio City Comprehensive Land Use Plan. Baguio City, Philippines.

Gonzales, M.A., 2009. More landslides feared in Baguio amid non-stop rains. *GMANews.TV*, 7 October 2009. http://www.gmanews.tv/story/174056/more-landslides-feared-in-baguio-amid-non-stop-rains (accessed 10 October 2009).

Jat, M.K., Garg, P.K., Khare, D., 2007. Monitoring and modeling urban sprawl using remote sensing and GIS techniques. *Int. J. Appl. Earth Observ. Geoinform.*, Vol.10, No.1, pp.26-43. http://linkinghub.elsevier.com/retrieve/pii/S0303243407000335 (accessed 8 February 2008).

Jennings, A.H., 1950. World's greatest observed point rainfalls. *American Meteorological Society.* Vol. 1, No. 70, pp. 4-5. http://docs.lib.noaa.gov/rescue/mwr/078/mwr-078-01-0004.pdf (accessed 20 November 2009).

Kawamura, J.G., 1971. Automatic recognition of changes in urban development from aerial photographs. Transactions on *Systems, Man, and Cybernetics*, Vol. 1, No. 3, pp. 230-239. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4308290 (accessed 20 July 2009).

Mananghaya, J., 2009. Northern Luzon towns isolated. *Philstar News Service*, 12 October 2009. http://www.philstar.com/Article.aspx?articleid=513366 (accessed 12 October 2009).

Mujtaba, S.M., 1994. Land use and environmental change due to urban sprawl: A remote sensing approach. New Delhi: Daya Publishing House.

MTDP, 2005 – 2010. Baguio City: Medium Term Development Plan 2005 - 2010. Baguio City, Philippines.

NSO, 2007. National Statistics Office: 2007 Census of Population. Manila, Philippines.

Orense, R.P., 2003. Geotechnical hazards-nature, assessment and mitigation. The University of the Philippines Press, Diliman, Quezon City, p.510.

Pepeng Report, 2009. Baguio City: Summary of Incidents Related to Tropical Storm Pepeng as of October 13, 2009. http://www.baguio.gov.ph/documents/pepengreport_oct13.pdf (accessed 10 November 2009).

Reddy, K.N., 1996. Urban redevelopment: a study of high rise buildings. http://books.google.com.ph (accessed 9 July 2009).

Refuerzo, A.P., 2009a. Creation of housing office urged. *News and Events*, 20 November 2009. http://www.baguio.gov.ph/index.php?q=content/creation-housing-office-urged (accessed 9 December 2009).

Refuerzo, A.P., 2009b. P1.5 million proposed for installation of screens to prevent flooding at city camp lagoon. *News and Events*, 20 November 2009. http://www.baguio.gov.ph/index.php?q=content/p15-million-proposed-installation-screens-prevent-flooding-city-camp-lagoon (accessed 9 December 2009).

Refuerzo, A.P., 2009c. Pepeng experience highlights need to address high risk areas, wholistic disaster management plan. *News and Events*, 19 October 2009. http://www.baguio.gov.ph/index.php?q=content/pepeng-experience-highlights-need-address-high-risk-areas-wholistic-disaster-management-plan (accessed 9 December 2009).

Sudhira, H.S., Ramachandra, T.V., Jagadish, K.S., 2004. Urban sprawl: metrics, dynamics and modeling using GIS. *Int. J. Appl. Earth Observ. Geoinform.*, Vol. 5, No. 1, pp. 26-39. http://linkinghub.elsevier.com/retrieve/pii/S0303243403000230 (accessed 11 November 2008).

Yeh, A.G.O., Li, X., 2001. Measurement and monitoring of urban sprawl in a rapidly growing region using entropy. *Photogr. Eng. Remote Sens.*, Vol. 67, No. 1, pp. 83-90. www.asprs.org/publications/pers/2001journal/january/2001_jan_83-90.pdf (accessed 11 November 2008).

ACKNOWLEDGEMENT

We thank the Engineering Research and Development for Technology and the Department of Science and Technology-Science Education Institute for funding this research. Obtaining the aerial photographs and satellite images for this study was facilitated by Engr. Maria Nicetas Caluya of F.F. Cruz, Inc., Maj. Dennis Alcarde of the Philippine Air Force, Lorna Cruz of J2, AFP. Various maps and reports were lent by Cordelia Lacsamana of the City Environment and Parks Management Office, Nellie Dela Cerna of the City Planning Office, Baguio City Hall. ENVI 4.3® and ArcGIS 9.3® were provided by the Department of Geodetic Engineering, University of the Philippines-Diliman, while the use of ERDAS-Imagine® was facilitated by NAMRIA thru Ms. Ofelia Castro.

AUTOMATED EXTRACTION OF PLANTATIONS FROM IKONOS SATELLITE IMAGERY USING A LEVEL SET BASED SEGMENTATION METHOD

K. Vogt*, B. Scheuermann, C. Becker, T. Büschenfeld, B. Rosenhahn, J. Ostermann

Institut für Informationsverarbeitung, Leibniz Universität Hannover, Appelstraße 9a, 30167 Hannover, Germany –
{vogt, scheuerm, becker, bfeld, rosenhahn, ostermann}@tnt.uni-hannover.de
http://www.tnt.uni-hannover.de/project/wipka/

KEY WORDS: Land Use, Vegetation, Segmentation, GIS, Automation

ABSTRACT:

In this article we present a method that extracts plantations from satellite imagery by finding and exploiting appropriate feature space projections. Segmentation is done using an automatic two-region segmentation based on the level set method. The behaviour of this algorithm is defined by a statistical region model that describes the similarity of regions using distances in arbitrary feature spaces. Subsequently different feature spaces will be evaluated regarding their plantation classification quality in an automatic fashion. The segmentation quality of our method is assessed by testing several orthophotos depicting a wide range of landscape types and comparing them with a manual segmentation. We show that a combination of simple texture based features like the structure tensor and the Hessian matrix are sufficient to facilitate an effective plantation segmentation scheme.

1. INTRODUCTION

When looking at land cover analysis, digital imagery is not only used for interpretation, but also for the verification and updating of tagged regions stored in GIS databases. Thus, the efficiency of today's workflow can be greatly improved by using automatic classification algorithms. In our knowledge-based interpretation and verification system GeoAIDA (Becker et al., 2008), several operators are used to hierarchically analyse image data. Area processing is done using a classification system which incorporates a Markov/Gibbs random field model, as proposed by Gimel'farb (Gimel'farb, 1996). Using this model, plantations (as seen in Figure 1) are particularly difficult to separate from general vegetation areas, since they may not notably differ in their statistical properties. To tackle this challenging task we propose a segmentation on a region model that works with a general feature space representation. We analyse several potential feature spaces, each constructed from the local region of a pixel. By analysing distances in these feature spaces, we can calculate the likelihood that the pixel belongs to a certain region of the current segmentation. For region detection we decided to use a state-of-the-art level set based segmentation framework that allows us to test our region model in a straightforward manner. The insights from the semi-automatic feature space comparison are then incorporated into the segmentation framework and finally tested on real orthophotos.

2. LEVEL SET SEGMENTATION

The variational approach for image segmentation used in our framework is based on the works of (Chan and Vese, 2001, Malladi et al., 1995, Paragios and Deriche, 1999, Rosenhahn et al., 2007). Using the level set formulation for the general problem of image segmentation has several advantages. To allow a convenient and sound interaction between constraints that are imposed on the contour itself and constraints that act on the two regions separated by the contour, the 1-D curve is embedded into a 2-D, image-like structure. Another important advantage of the level set representation is the natural given possibility to handle topological changes of the 1-D curve. This makes it particularly useful for plantation segmentation, where no reliable constraints for the region shapes can be formulated.

* Corresponding author.

In the case of a two-region segmentation, the level set function $\varphi : \Omega \rightarrow \mathbb{R}$ splits the image domain Ω into the two regions $\Omega_1, \Omega_2 \subseteq \Omega$ with

$$\varphi(x) = \begin{cases} \geq 0, \text{ if } x \in \Omega_1 \\ < 0, \text{ if } x \in \Omega_2 \end{cases} . \quad (1)$$

The boundary between the object that is to be extracted and the background is represented by the zero-level line of the function φ. Like most of the works on level set segmentation, we will focus on this special segmentation case with two regions. In our work these will represent the object classes plantations and non-plantations. As shown in (Chan and Vese, 2001), we will require an optimal contour to minimise the so called *Chan-Vese energy functional* that is:

$$E(\varphi) = -\int_\Omega \Big(H(\varphi) \log p_1 + (1 - H(\varphi)) \log p_2 \Big)\, d\Omega + \nu \int_\Omega |\nabla H(\varphi)|\, d\Omega \quad (2)$$

where $\nu \geq 0$ is a weighting parameter between the two given constraints, p_i are probabilities and $H(s)$ is a regularised Heaviside function. The regularised Heaviside function is needed to build the Euler-Lagrange equation and to make it possible to indicate which region a pixel belongs to at each iteration step. Minimising the first term maximises the total a-posteriori probability given the the two probabilities p_1 and p_2 of Ω_1 and Ω_2, i.e., pixels are assigned to the most probable region according to the Bayes rule. The second term minimises the length of the contour and acts as a smoothing term.

Minimisation of the Chan-Vese energy functional (2) can be performed by solving the corresponding Euler-Lagrange equation to φ

$$\frac{\partial \varphi}{\partial t} = \delta(\varphi) \left(\log \frac{p_1}{p_2} + \nu \operatorname{div} \left(\frac{\nabla \varphi}{|\nabla \varphi|} \right) \right) , \quad (3)$$

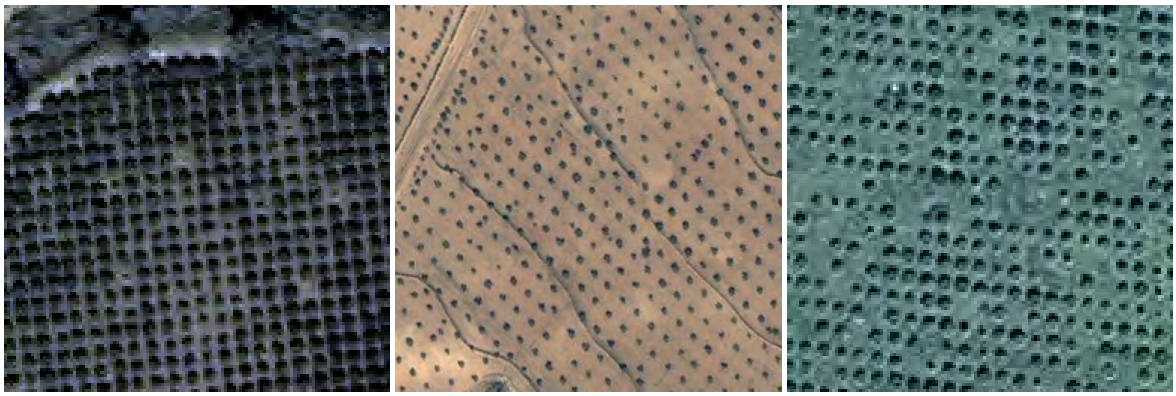


Figure 1: Plantation samples from different locations

where $\delta(s)$ is the derivative of $H(s)$, with respect to its argument. Starting with some initial contour φ^0 and given the probabilities p_1 and p_2 one has to solve the following initial value problem:

$$\begin{cases} \varphi(x,0) = \varphi^0 \text{ for } x \in \Omega \\ \dfrac{\partial \varphi}{\partial t} = \delta(\varphi)\left(\log\dfrac{p_1}{p_2} + \nu \operatorname{div}\left(\dfrac{\nabla\varphi}{|\nabla\varphi|}\right)\right) \end{cases} . \quad (4)$$

Figure 2 shows the workflow of our segmentation algorithm. The way the two probabilities p_1 and p_2 are modelled is a very important factor for the quality of the segmentation process. The next section will, therefore, introduce a statistical region model capable of estimating probability densities in arbitrary feature spaces.

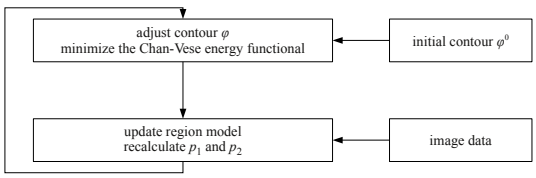


Figure 2: Workflow of the segmentation algorithm

3. STATISTICAL REGION MODEL

The purpose of the region model is to provide a model of the statistical properties underlying the region that is to be extracted by our segmentation algorithm. The interface between the region model and segmentation algorithm is defined by a function $p(x, y)$, mapping each pixel position to the probability that this pixel belongs to the modelled region of the current segmentation. Since $p(x, y)$ depends on the current contour of the region, which will change during the progression of the segmentation, a good presegmentation is vital in order to achieve high quality results. To estimate $p(x, y)$, we use the function $f(x, y)$, which defines a projection from image to feature space. We can now easily calculate the mean distance between each position and the modelled region in an arbitrary feature space. The probability function $p(x, y)$ is now defined by scaling the resulting distances to the range $[0; 1]$, while considering that larger distances will map to a lower probability and vice versa.

In case that several functions $f_c(x, y)$ are available, each will provide one independent feature channel c. The function $p(x, y)$ can then be calculated by multiplying the individual probability functions $p_c(x, y)$.

3.1 Distance Function

When incorporating several feature channels, a consistent scaling of the feature distances might become quite complex. We will, therefore, present a distance function based on the normalised cross correlation (NCC). The feature vectors $i, j \in \mathbb{R}^n$ will be interpreted as discrete signals of length n. The NCC is then defined as

$$\varrho_{ij} = \frac{1}{n-1}\sum_{t=0}^{n-1}\frac{(i(t)-\bar{i})\,(j(t)-\bar{j})}{\sigma_i \sigma_j}, \quad (5)$$

whereby $\bar{i}$ and $\bar{j}$ represent the mean component values of i and j. Since the range of the NCC is known, scaling the distances is quite trivial. Handling of anti-correlations depends on the application so that ϱ_{ij} may either be mirrored around or clamped at 0, keeping all values positive.

$$d_{NCC_1}(i,j) = 1 - |\varrho_{ij}| \quad (6)$$

$$d_{NCC_2}(i,j) = 1 - [\varrho_{ij}]^+, \quad (7)$$

$$with\ [x]^+ = \begin{cases} 0 & , x < 0 \\ x & , else \end{cases}$$

3.2 Presegmentation

In most cases, a good presegmentation can be derived from existing geodata. If the data is not available or too outdated, there is an easy method for generating one by analysing the frequency spectrum of image blocks from the source image. Typical images of plantations show a peak in a certain frequency range. After an appropriate range was learned from existing image data (see section 3.3.1), each image block is tested for the occurrence of such a peak and assigned to either the inner or outer region of the presegmentation. Figure 3 shows the result of the presegmentation step for one test case.

3.3 Feature Selection

Finding appropriate feature spaces is no easy task and is often done manually. Instead of designing one that is optimised for plantation segmentation, our strategy is to generate a wide array of generic feature spaces and compare them in a systematic manner. This selection process is guided by a list of training sets and can also be used to find feature spaces for other object classes. The general workflow shown in figure 4 is further explained in the following sections.

Figure 3: The presegmentation mask for a sample scene

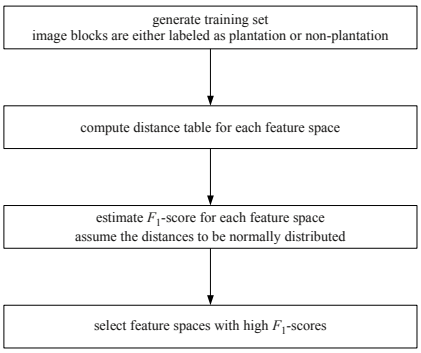


Figure 4: Workflow of the feature selection

3.3.1 Training Sets. In total, eight training sets were produced, each of them chosen to represent a different type of geographical location. Together, the training sets cover an area of approximately $75km^2$ subdivided into 4400 image blocks. These image blocks are then manually classified as being either of the plantation or non-plantation type.

For each training set, a distance table can be generated. A distance table contains the distances between all image blocks for a given feature space. It is convenient to sort the entries so that all plantation and non-plantation blocks are grouped together. The table can now be subdivided into four quadrants:

- Intra-distances between non-plantations
- $2\times$ Inter-distances
- Intra-distances between plantations

A distance table from a small training set is shown in Figure 5.

3.3.2 Classification Model. Our goal is to find feature-spaces that produce small intra-distances and large inter-distances. Of course, *small* and *large* are entirely subjective terms. It is sufficient that intra- and inter-distances are separable. To that end, we will treat that task as a classical binary classification problem.

Using the distance tables we can find normal distributions describing the intra- and inter-distances. As shown in Figure 6, by analysing the overlap of the normal distributions, we can compute the true positive (TP), false positive (FP), true negative (TN) and false negative (FN) rates. It is obvious that features with minimal overlaps will generate better classification results. To get a better understanding of the classification capacity of a given feature space, we will also calculate precision, recall and the corresponding F_1-score (Van Rijsbergen, 1979):

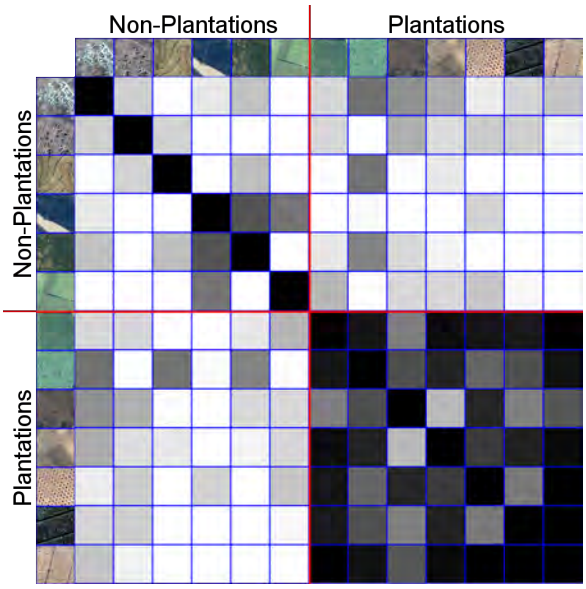


Figure 5: The distance table of a feature contains the distances between all image blocks in the training set. Dark entries signify a small distance in a given feature space and therefore a high similarity.

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (10)$$

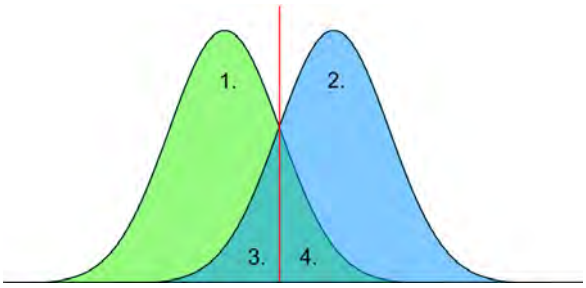


Figure 6: Intra-distances (green) and inter-distances (blue) are approximated by normal distributions. The numbered areas represent: 1. True Positive 2. True Negative 3. False Negative 4. False Positive

3.3.3 Tested Features. Features are generated in a two-step process. First, the image blocks are transformed during a preprocessing step. The resulting image blocks are then projected into a feature space of arbitrary dimension. The following lists present the image and feature space transformations that were tested:

Image transformations: Fourier transformation, structure tensor, Hessian matrix, local entropy, local fractal dimension, normalised difference vegetation index (*NDVI*)

Feature transformations: Histogram, co-occurrence matrix, autocorrelation, frequency spectrum

Some of our features were from the literature (Pratt, 2001), (Brox et al., 2006), (Soille and Rivest, 1996) and (Rouse et al., 1974).

Table 7 contains the results of the different transformation combinations in terms of their F_1-scores. Higher scores indicate better

	Histogram	Co-occurrence matrix	Autocorrelation	Frequency spectrum
Identity	0.78	0.76	0.61	0.75
Fourier transformation	0.83	0.79	0.78	0.73
Local entropy	0.68	0.67	0.52	0.71
Local fractal dimension	0.54	0.53	0.51	0.57
NDVI	0.76	0.53	0.59	0.75
Structure tensor	0.83	0.72	0.61	0.78
Hessian matrix	0.89	0.89	0.58	0.82

Table 7: Median of the F_1-scores across the training sets. Best results are emphasised.

classification results, whereby a 1.0 would imply a perfect classification. Overall, there is a significant variability between each training set, but texture based features generally generate better results. Surprisingly, the NDVI feature only works well on half the training sets, while autocorrelation results tend to be situated at the bottom of the rankings, despite the strikingly regular structure usually displayed by plantations.

4. RESULTS

Input scenes for the training sets and the plantation extraction tests were selected from IKONOS satellite images with a 1m resolution in panchromatic mode. Great care was exercised to select scenes that feature a wide variety of landscape types, and especially most plantation subtypes.

Table 8 shows the segmentation results for five test scenes, using different feature space combinations. The feature spaces used for the segmentations in Figure 9 - 11 were selected according to the results of the feature selection step in table 7. As such, a combination of three feature spaces was used:

- *Histogram from Hessian matrix*
- *Histogram from structure tensor*
- *Histogram from Fourier transformation*

The left image shows a manually generated reference mask, while the right image shows the result from our segmentation algorithm. Overall, our algorithm produces excellent recall, but only mediocre sensitivity results, which were due to plantation regions heavily expanding into background regions. This effect is emphasised if the background region contains object classes that do not appear homogeneous in the chosen feature space.

5. CONCLUSIONS AND FUTURE WORK

Our results have shown that feature spaces well suited for the extraction of specific object classes can be selected in an automatic manner. For some scenes, a segmentation quality of up to 0.94 (F_1-score) was reached. Still, the choice of a two-region segmentation scheme proved to be too restricting, since non-plantation object classes do not appear homogeneous in any single feature space that was tested. Our plan is, therefore, to update our framework to support a multi-region segmentation algorithm, allowing us to define and extract multiple object classes from a scene at the same time. Simultaneously, the catalogue of evaluated feature spaces will be expanded to get a better understanding of which feature space encapsulates the statistical properties of each object class best.

ACKNOWLEDGEMENTS

The work is funded by BKG, Frankfurt (Main) Germany. We gratefully acknowledge this support.

REFERENCES

Becker, C., Ziems, M., Büschenfeld, T., Heipke, C., Müller, S., Ostermann, J. and Pahl, M., 2008. Multi-hierarchical quality assessment of geospatial data. ISPRS.

Brox, T., Weickert, J., Burgeth, B. and Mrazek, P., 2006. Nonlinear structure tensors. Image and Vision Computing 24(1), pp. 41–55.

Chan, T. and Vese, L., 2001. Active contours without edges. IEEE Transactions on Image Processing 10(2), pp. 266–277.

Gimel'farb, G. L., 1996. Texture modeling by multiple pairwise pixel interactions. IEEE Trans. Pattern Anal. Mach. Intell. 18(11), pp. 1110–1114.

Malladi, R., Sethian, J. and Vemuri, B., 1995. Shape modelling with front propagation: A level set approach. IEEE Transaction on Pattern Analysis and Machine Intelligence 17(2), pp. 158–174.

Paragios, N. and Deriche, R., 1999. Unifying boundary and region based information for geodesic active tracking. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, IEEE Computer Society Press, Forth Collins, Colorado, pp. 300–305.

Pratt, W. K., 2001. Digital Image Processing: PIKS Inside. John Wiley & Sons, Inc., New York, NY, USA.

Rosenhahn, B., Brox, T. and Weickert, J., 2007. Three-dimensional shape knowledge for joint image segmentation and pose tracking. International Journal of Computer Vision 73(3), pp. 243–262.

Rouse, W., J., H., R., Haas, A., J., Shell, D.W., Deering, J.C. and Harlan, 1974. Monitoring the vernal advancement of retrogradation of natural vegetation. Final report, Greenbelt, MD: NASA/GSFC.

Soille, P. and Rivest, J.-F., 1996. On the validity of fractal dimension measurements in image analysis. Journal of Visual Communication and Image Representation 7(3), pp. 217–229.

Van Rijsbergen, C. J., 1979. Information Retrieval, 2nd edition. Dept. of Computer Science, University of Glasgow.

	Scene 1			Scene 2			Scene 3			Scene 4			Scene 5		
	Recall	Precision	F_1-score	Recall	Precision	F_1-score	Recall	Precision	F_1-score	Recall	Precision	F_1-score	Recall	Precision	F_1-score
1. Only presegmentation	0.85	0.65	0.74	0.20	0.97	0.33	0.46	0.65	0.54	0.88	0.59	0.70	0.92	0.59	0.72
2. Histogram from structure tensor	0.98	0.68	0.80	0.28	0.79	0.41	0.62	0.64	0.63	0.91	0.75	0.82	0.82	0.64	0.72
3. Histogram from Hessian matrix	0.98	0.69	0.81	0.90	0.82	0.86	0.84	0.61	0.71	0.96	0.66	0.78	0.92	0.60	0.73
4. Frequency spectrum from image	0.85	0.65	0.74	0.20	0.97	0.33	0.46	0.65	0.54	0.88	0.59	0.70	0.92	0.59	0.72
5. Histogram from NDVI	0.77	0.65	0.70	0.92	0.87	0.89	0.96	0.62	0.75	0.84	0.50	0.63	0.94	0.52	0.67
6. Feature combination (2) + (3)	0.98	0.68	0.80	0.91	0.80	0.85	0.73	0.64	0.69	0.93	0.72	0.81	0.89	0.63	0.74
7. Feature combination (2) + (4)	0.98	0.68	0.80	0.28	0.77	0.41	0.62	0.64	0.63	0.91	0.75	0.82	0.82	0.64	0.72
8. Feature combination (2) + (5)	0.97	0.75	0.85	0.93	0.87	0.90	0.96	0.62	0.75	0.85	0.50	0.63	0.94	0.52	0.67
9. Feature combination (3) + (4)	0.98	0.70	0.82	0.90	0.82	0.85	0.84	0.61	0.71	0.96	0.66	0.78	0.92	0.60	0.73
10. Feature combination (3) + (5)	0.94	0.76	0.84	0.92	0.87	0.89	0.96	0.62	0.75	0.85	0.50	0.63	0.94	0.52	0.67
11. Feature combination (4) + (5)	0.77	0.65	0.70	0.93	0.87	0.90	0.96	0.62	0.75	0.84	0.50	0.63	0.94	0.52	0.67
12. Feature combination (2) + (3) + (4)	0.98	0.68	0.80	0.91	0.80	0.85	0.73	0.64	0.69	0.93	0.72	0.81	0.89	0.63	0.74
13. Feature combination (2) + (3) + (5)	0.98	0.74	0.85	0.92	0.87	0.90	0.96	0.62	0.76	0.85	0.50	0.63	0.94	0.52	0.67
14. Feature combination (2) + (4) + (5)	0.97	0.75	0.85	0.93	0.87	0.90	0.96	0.62	0.75	0.85	0.50	0.63	0.94	0.52	0.67
15. Feature combination (3) + (4) + (5)	0.94	0.76	0.84	0.92	0.87	0.89	0.96	0.62	0.76	0.85	0.50	0.63	0.94	0.52	0.67
16. Feature combination (2) + (3) + (4) + (5)	0.98	0.75	0.85	0.92	0.87	0.90	0.96	0.62	0.76	0.85	0.50	0.63	0.94	0.52	0.67

Table 8: Segmentation results for five test scenes using different feature space combinations. Best results for each scene are emphasised.

Figure 9: Manual reference segmentation (left) and our fully automatic segmentation (right)

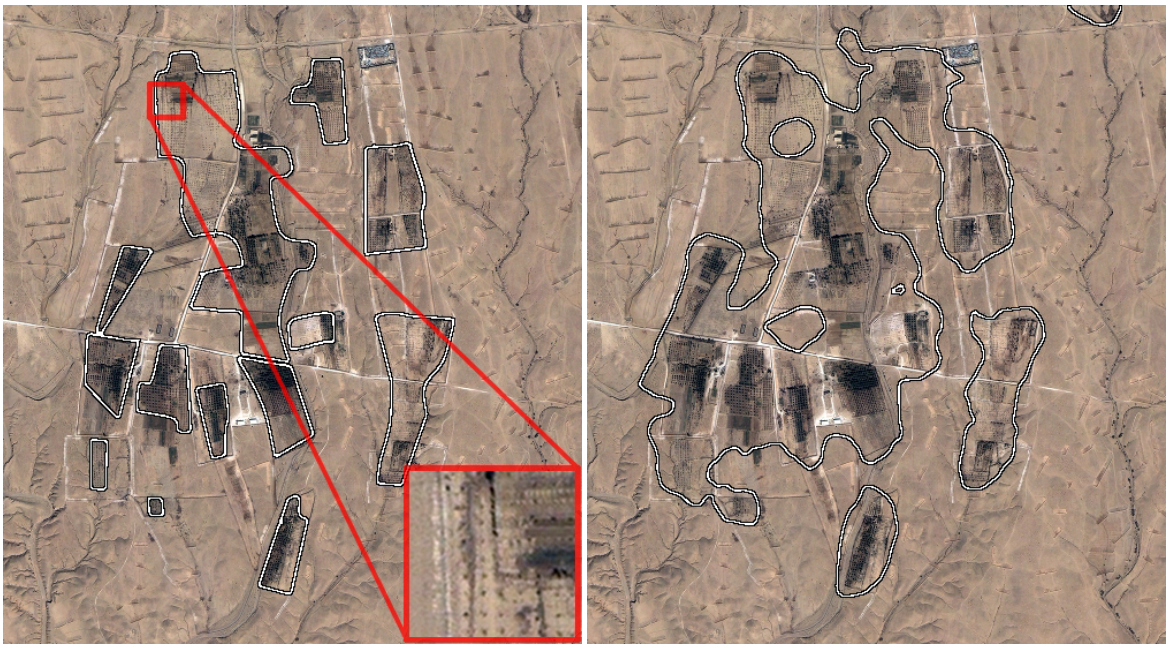


Figure 10: Manual reference segmentation (left) and our fully automatic segmentation (right)

Figure 11: Manual reference segmentation (left) and our fully automatic segmentation (right)

RETRIEVAL OF BIOPHYSICAL VEGETATION PRODUCTS FROM RAPIDEYE IMAGERY

F. Vuolo [a, *], C. Atzberger [b], K. Richter [c], G. D'Urso [d], J. Dash [a]

[a] School of Geography, University of Southampton, Highfield, Southampton SO17 1BJ, UK;
[b] Joint Research Centre of the European Commission, JRC, Institute for the Protection and Security of the Citizen, MARS unit, Via Enrico Fermi 2749, 21027 Ispra (VA), Italy;
[c] Department of Geography, Faculty of Geosciences, Ludwig-Maximilians-University Munich, Luisenstr. 37, 80333 Munich, Germany;
[d] Department of Agricultural Engineering and Agronomy (DIAAT), University of Naples "Federico II", via Università 100, 80055 Portici (Na), Italy.

KEYWORDS: Agriculture, Vegetation, Monitoring, Retrieval, Multispectral.

ABSTRACT:

The accurate estimation of canopy biophysical variables at sufficiently high spatial and temporal resolutions is a key requirement for operational applications in the agricultural sector. In this study, recently available multispectral RapidEye sensor data were tested for their operational suitability to estimate canopy biophysical variables in the Italian Campania region. For this purpose, two model inversion methods and two commonly used vegetation indices were applied to estimate leaf area index (LAI), canopy chlorophyll content (CCC) and leaf chlorophyll content (LCC) from a range of crops. The physically based approaches outperformed the empirical methods, with a slightly higher retrieval accuracy of the look-up table (LUT) than of the neural network (NN) approach. However, the NN method performs much faster, rendering it potentially more appropriate for application in large areas. The empirical models showed dependencies of sensor and crops, but still performed reasonable in the estimation of LAI and CCC. Results demonstrated the suitability of RapidEye sensor data to retrieve canopy biophysical variables of agricultural areas.

1. INTRODUCTION

1.1 Crop monitoring

The regular and accurate mapping of crop status is an important requirement for a sustainable agricultural management. It enables, for instance, the early detection of crop water stress or nitrogen deficiencies, thus helping farmers to mitigate potential crop damages while reducing environmental impacts. For this purpose, frequent information of crop status at sufficiently high spatial resolutions is required, being of particular importance for heterogeneous agricultural regions, characteristic for Southern Italy. Remotely sensed data from air- or space-borne platforms offer an interesting alternative to cost and labour intensive ground measurements. Earth observation (EO) data with improved spatial and temporal resolutions, such as those from the RapidEye constellation (http://www.rapideye.de/), offer new opportunities for a sustainable agricultural management.

1.2 Biophysical variables and retrieval techniques

In the present study, three of the key biophysical variables of interest for precision farming applications were examined: leaf area index (LAI) (e.g., Moran et al., 1995), leaf chlorophyll content (LCC) and canopy chlorophyll content (CCC) (e.g., Baret et al., 2007). LAI, a key variable of vegetation, characterizes the leaf surface available for energy and mass exchange between surface and atmosphere (Moran et al., 1995). Different definitions of LAI have been used in the literature depending on vegetation type and measurements (Jonckheere et al., 2004), such as green LAI ('GLAI', e.g. Migdall et al., 2009), effective LAI ('L_e', Chen and Black, 1992) or plant area index ('PAI', Neumann et al., 1989).

Chlorophyll content was found to be directly related to nitrogen (N) availability of the leaves (e.g. Evans, 1989). Therefore, the sensitivity of the solar reflective domain to chlorophyll content is usually used to quantify the plant nitrogen status. Baret et al. (2007) demonstrated that the relationship between canopy chlorophyll content and N is more robust over years and development stages than the correlation at leaf level. Thus, canopy chlorophyll content presents greater potential than leaf chlorophyll content to detect vegetation stress and should be the privileged variable to be retrieved.

A variety of methods have been proposed to estimate these biophysical variables from remotely sensed data (Baret and Buis, 2008). The majority of the studies have used (semi-) empirical relationships between the biophysical variables of interest and a combination of spectral bands, namely vegetation indices (VI). These methods, successfully applied to a number of applications (Glenn et al., 2008), are fast and easily implementable at large data sets and thus suitable for operational purposes. The Weighted Difference Vegetation Index (WDVI) (Clevers, 1989), for instance, is being used for operational retrievals of LAI in the context of Irrigation Advisory Services in Southern Italy (De Michele et al., 2009). Moreover, information on canopy chlorophyll content is being routinely distributed to users through the MERIS Terrestrial Chlorophyll Index (MTCI) (Dash and Curran, 2007). Currently, MTCI is operationally available only at medium spatial resolution, but it will be supplied by future ESA's Sentinel-2 optical system at finer spatial resolution.
Despite the wide use of these approaches, VIs are limited in their global estimation performance since calibration is mostly required to account for changing conditions. This includes for instance differences in sensor types and crop canopy architecture, changing illumination and viewing geometries or

* Corresponding author

varying soil backgrounds (Colombo et al., 2003). With the advancement in developing radiative transfer models (RTM), these aspects can be considered by means of physical principles. Therefore, new perspectives have opened up for reliable and accurate estimations of biophysical products in the context of operational applications (Bacour et al., 2006).
However, these models have also limitations, such as the need for parameterization and high computational demand. Furthermore, the ill-posed inverse problem must be considered: different parameter combinations may produce almost identical spectra, resulting in significant uncertainties in the estimation of biophysical vegetation variables (Atzberger, 2004). Even though this problem affects as well empirical approaches, it is often only discussed in the context of RTM model inversion.

The objective of this study is to evaluate the performance of RapidEye sensor data to estimate LAI, leaf chlorophyll content and canopy chlorophyll content. Two model inversion techniques, i.e. look-up tables (LUT) and neural networks (NN), are applied for this purpose. In order to evaluate the performance of current operational VIs, the analysis is extended to the estimations of LAI and canopy chlorophyll content by using pre-calibrated equations based on WDVI and MTCI. The suitability of the data and the retrieval techniques are discussed.

2. MATERIAL AND METHODS

2.1 Field Campaign

The data used in this study are based on satellite acquisitions and a ground measurement campaign at the "Piana del Sele" study site (Lat. 40.52 N, Long. 15.00 E), which is one of the largest agricultural areas of the Italian Campania region, Southern Italy. The area is characterized by irrigated agriculture (mainly forages and fruit trees) with an average field size of about 2 hectares (De Michele et al., 2009).
A total number of 36 LAI and leaf chlorophyll measurements were acquired simultaneously at different sites and for a range of crops, including fruit trees (such as peach and apricot), maize, cereals and different vegetables. LAI measurements were carried out by means of the Plant Canopy Analyzer LAI-2000 instrument (LICOR Inc., Lincoln, NE, USA). Due to its measurement principle, the sensor does not distinguish photosynthetically active leaf tissue from other plant elements, such as stems, flowers or senescent leaves. Moreover, the clumping effect, i.e. non-random positioning of canopy elements, is neglected. Thus, the here used term 'LAI' stands for effective PAI ('PAI_e') (Darvishzadeh et al., 2008).
Measurements were performed in order to cover an Elementary Surface Unit (ESU) of approximately 400 m^2 geolocated by means of a GPS device (accuracy 3-5 m). The average value of LAI, resulting from a set of 20 above and below canopy readings, was considered to be representative for the respective ESU. The standard deviations of the measurements were kept as a measure of uncertainty.
Leaf chlorophyll content was measured by using a SPAD-502 Leaf Chlorophyll Meter (MINOLTA, Inc.). At each ESU, 30 measurements of leaves in different layers were randomly performed and averaged to a final representative value. Crop specific calibration functions (peach tree: Marchi et al., 2005; maize: Haboudane et al., 2001; other crops: SPARC, 2004) were applied to convert the SPAD values into leaf chlorophyll content [$\mu m\ cm^{-2}$]. The total canopy chlorophyll content [$g\ m^{-2}$] was finally obtained by multiplying leaf chlorophyll content with the corresponding LAI of each ESU.

2.2 Remote Sensing data

Multispectral remote sensing data from RapidEye sensor were acquired on 17th August 2009 (at 10:35 UTC). This recently launched constellation (August 2008) of five identical EO satellites records radiance in five broad bands corresponding to blue, green, red, red-edge and near-infrared (NIR) part of the electromagnetic spectrum. The sensors provide a spatial resolution of 5 m and are therefore potentially very suitable for agricultural applications.
Four images, acquired within a few seconds, with a maximum across-track incident angle of 5° were adequate to cover the study site (about 560 km^2). Radiometrically calibrated Level 3A data were provided with a geometric accuracy of 13.95 m (root mean square error, RMSE = 6.50 m). Further geometric correction was performed using Ground Control Points (GCPs), resulting in a final geolocation accuracy of about 3 m.
The first image tile was atmospherically corrected by using ATCOR-2/3 (Richter, 1998). The spectral reflectance of known reference targets (i.e., asphalt, sea water, concrete and sand) was used for the retrieval of atmospheric properties. Subsequently, an empirical line method was applied to correct the other three images. For this purpose, uniform areas in the overlapping regions between adjacent images were considered: twenty zones of about 200 m^2 representing dark and bright surfaces were selected for each image and correction functions were derived for each spectral band.
To account for the accuracy of geometric correction and ground biophysical variable measurements, the final mosaicked image was resampled to a spatial resolution of 15 m.

2.3 Radiative Transfer Modelling

The well-known and widely used coupled PROSPECT+SAILH model ('PROSAIL', Jacquemoud et al., 2009) was chosen for the study. PROSAIL is a combination of the leaf model PROSPECT–4 (Feret et al., 2008) and the canopy model SAILH (Verhoef 1984, 1985; Kuusk 1991). It calculates the bi-directional reflectance of homogeneous canopies as a function of several structural and biophysical variables (see Table 1), soil reflectance, illumination and viewing geometry.

2.3.1 Model inversion with look-up tables (LUT)

Even though it is a relatively simple method, the look-up table (LUT) approach is one of the most robust and accurate model inversion strategies. It has been applied in combination with the PROSAIL model by a number of studies (e.g. Darvishzadeh et al., 2008; Richter et al., 2009; Weiss et al., 2000), successfully retrieving biophysical variables of different crop types and at different sites.
To set up the inversion, a synthetic data base was established with the PROSAIL model simulating RapidEye spectral band configuration using the specific band sensitivity functions. A LUT size of 100000 different combinations of variables was chosen according to Weiss et al. (2000). The variables and model parameters were randomly sampled using uniform distribution laws and according to typical ranges found in the literature for agricultural land use (Table 1). Model inversion was performed using a simple cost function calculating the RMSE between measured and simulated spectra. The solution was regarded as the average of the variable combinations found within less than 20 % of the lowest RMSE value (e.g. Richter et al., 2009).

2.3.2. Model inversion with neural networks (NN)

Neural networks were included in this study as an alternative mean to (rapidly) invert RTM over large areas. The synthetic data base generated for the LUT approach was used to train the network. Training permits a net to learn the intrinsic relation between some input variables (here the canopy reflectance spectra) and one or more output variables (here the sought biophysical variables). Setting up the network structure and network training may be a time consuming process. However, once trained, the sought biophysical variables can be retrieved immediately.

To prevent overfitting and overspecialisation several measures were taken. First, the network was kept compact using a single hidden layer with only five neurons. Second, three variables were modelled at the same time to avoid over-specialisation: LAI, leaf chlorophyll content and soil reflectance scaling factor (α_{soil}). Finally, the early stopping technique was applied to further improve network generalization. For this purpose, the patterns generated with PROSAIL were divided into two subsets. The first subset (75 % of the pattern) was used for updating the weights and biases of the network (training dataset). The error on the test dataset (the remaining 25 %) was monitored during the training process. The training was stopped automatically when the error in the test dataset started to rise as this indicates network overfitting.

Model Variables		Units	Range
PROSPECT			
N	Leaf structure index	unitless	1.3-2.0
C_{ab}	Leaf chlorophyll content	[µg cm^{-2}]	10-70
C_m	Leaf dry matter content	[g cm^{-2}]	0.004-0.007
SAILH:			
LAI	Leaf area index	[m^2 m^{-2}]	0-6
ALA	Average leaf angle	[degree]	35-70
HotS	Hot spot parameter	[m m^{-1}]	0.01-1
α_{soil}	Soil reflectance scaling factor	unitless	0.6-1.4
θ_s	Sun zenith angle	[degree]	28
θ_v	View zenith angle	[degree]	5
ϕ	Sun – sensor azimuth angle	[degree]	71

Table 1. Range of model input variables used to establish the synthetic canopy reflectance data base for NN and LUT based model inversions.

2.3.3. Empirical model: vegetation indices (VI)

The WDVI is based on the reflectance in the NIR and red wavelength ranges. Calculation of WDVI requires information of the soil line slope, which can be directly derived from the imagery. A logarithmic relationship was used to estimate LAI from WDVI, which was calibrated during several field campaigns in the study site in the last years (R^2=0.64) (D'Urso and Belmonte, 2006).

The MTCI (Dash and Curran, 2007) was calculated from NIR, red edge and red spectral bands. A linear equation calibrated using ground data (R^2=0.80) (Dash at al., 2010) was adopted in this study to estimate the canopy chlorophyll content. Detailed description of the indices can be found in D'Urso and Belmonte (2006) and Dash and Curran (2007), respectively.

3. RESULTS AND DISCUSSION

3.1 Retrieval of leaf and canopy variables

Estimations of LAI using the two inversion methods performed well with a slightly higher accuracy from the LUT (RMSE=0.64; R^2=0.76) than from the NN method (RMSE=0.72; R^2=0.71). With the WDVI, a lower estimation accuracy was achieved (RMSE=1.14; R^2=0.57). Measured against simulated LAI values are presented in Fig.1a-c.

For canopy chlorophyll content, a high retrieval accuracy was obtained from the LUT (RMSE=0.39 g m^{-2} and R^2= 0.78) and a slightly lower from the NN (RMSE=0.43 g m^{-2} and R^2=0.74). Application of the MTCI achieved a lower accuracy than the physically based approaches (RMSE=0.86 g m^{-2} and R^2=0.73). Correlations between estimated and measured canopy chlorophyll content values are shown in Fig.1d-f.

In case of leaf chlorophyll content, all approaches failed to give reliable estimates: by the LUT a RMSE of 15.1 µg cm^{-2} and by the NN a RMSE of 11.3 µg cm^{-2} was achieved (not shown).

Regarding crop specific differences in retrieval accuracy (Table 3), LAI values were generally estimated best for fruit trees. Estimation uncertainties may be explained by the non-linear relationship between reflectance and LAI, leading to saturation at higher LAI values, as visible in Fig. 1 for all approaches. Moreover, a possible presence of clumped leaves may strengthen the underestimation of higher LAI values (i.e., LAI > 3), especially in case of maize.

Canopy chlorophyll content was obtained with a reasonable accuracy for maize and partly fruit trees using the model inversion techniques.

The overall poor retrieval accuracy of leaf chlorophyll content is also reflected in the crop specific RMSE. Only for maize, the LUT achieved reasonable results with RMSE of 5.9 µg m^{-2}.

The retrieval accuracy of leaf characteristics from canopy spectra depends on the strength of the signal transmitted from leaf to canopy level, which is mainly controlled by structural variables such as LAI or leaf angle (Asner, 1998). Thus, compensations between LAI and leaf chlorophyll content may occur, leading to the well-known ill-posed inverse problem (Combal et al., 2002). Strong improvements in the estimation accuracy were also observed in other studies when using the product between these two variables (Baret et al., 2007).

A further explanation for the poor estimation of leaf chlorophyll content may be the presence of heterogeneous canopies (such as fruit trees and maize), not corresponding to the turbid medium assumption of the used model. The application of more complex models, such as GeoSAIL proposed by Huemmrich (2001), might improve the retrieval performance (Richter et al., 2009). A further improvement may result from object-based inversion strategies (Atzberger, 2004).

The lower retrieval accuracy of the VI confirms the need of sensor-specific calibration, in particular for the MTCI, which was specifically designed for narrow visible/NIR wavebands. The red edge band originally used to calculate the MTCI is based on MERIS spectral band characteristics with a spectral bandwidth of about 10 nm (703.75 - 713.75 nm). RapidEye red edge band is instead sampled in a spectral bandwidth of 40 nm (690 - 730 nm). Therefore, adaptation of the equation to broad band spectral characteristics, as for RapidEye sensors, might be required. Further investigation is needed for this issue.

Results obtained with WDVI confirm that a sensor and crop specific calibration is required.

A map of spatially distributed canopy chlorophyll content (using the NN) is depicted in Fig. 4, representing the area of "Piana del Sele" in the Italian Campania region. Such maps are possible inputs in the context of precision farming applications.

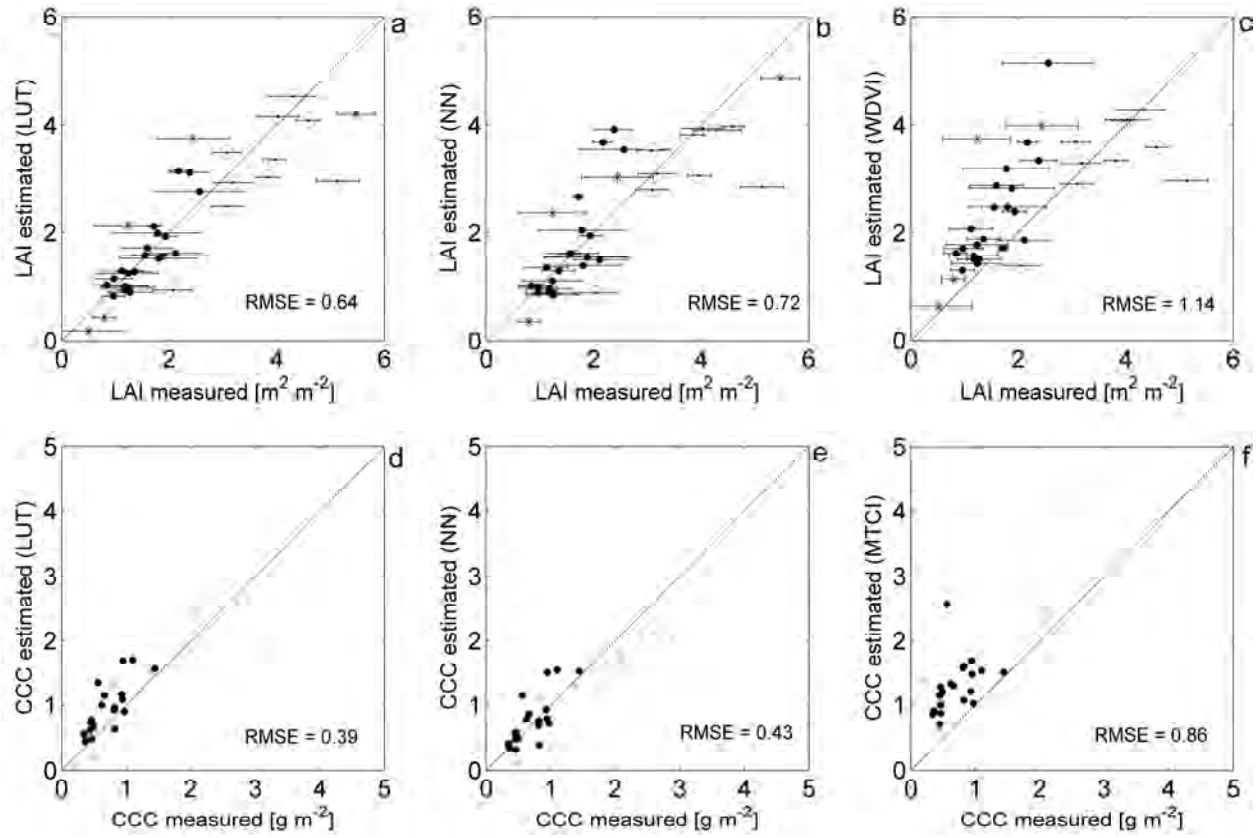


Figure 2. Estimated versus measured biophysical variables of different crops (Piana del Sele, Italian Campania region). 1a: LAI with LUT approach, 1b: LAI with NN approach, 1c: LAI from WDVI, 1d: canopy chlorophyll content (CCC) from LUT, 1e: CCC estimated with NN, 1f: CCC from MTCI. Symbols correspond to: '·' fruit trees, 'o' maize and '*' other crops. Error bars in 1a-c indicate standard deviations of the LAI measurement.

Crop type	LAI [m^2 m^{-2}]			CCC [g m^{-2}]			LCC [µg m^{-2}]	
	LUT	NN	VI	LUT	NN	VI	LUT	NN
Fruit trees(1)	0.35	0.61	0.95	0.34	0.25	0.7	18.3	8.7
Maize	0.89	0.91	0.82	0.50	0.70	1.18	5.9	10.4
Others (2)	0.93	0.72	2.03	0.31	0.28	0.7	13.1	19.4
Combined	0.64	0.72	1.14	0.39	0.43	0.86	15.1	11.3

(1) includes peach, apricot, kiwi and plum trees
(2) includes aubergines, alfalfa, pepper, artichokes and cereal

Table 3. Crop specific (and combined) RMSE between measured and estimated biophysical variables using LUT, NN and the two VI approaches (MTCI for CCC and WDVI for LAI estimation).

3.2 Operational suitability

An important issue for the use of physically based retrieval techniques in the context of operational applications is the time required for inverting RTM over large areas. Both inversion methods perform rather fast in comparison to traditional approaches, such as iterative optimisation techniques. However, the NN method outperforms clearly the LUT in this regard.

In pixel-based inversions, redundant LUT searches are being performed since many signatures are similar. Therefore, in order to render the LUT inversion procedure more effective and faster, an unsupervised classification was applied to the imagery before further processing, grouping the reflectance spectra into a certain number of classes. This number depends on the heterogeneity of the region and sensitivity analyses must be carried out to obtain the optimal number of classes, reducing redundancy without losing important spectral information. For the study area, 2000 classes were chosen, reducing the computational load almost 850 times (original number of pixel 1.7 mil.). The ISODATA clustering method of the Erdas Imagine software, which uses minimum spectral distance formula, was applied. Maximum number of iterations was set to '6' and the convergence threshold to '0.95'.

Each of the 2000 input spectra was calculated as the average of all spectra contained in one class. Processing of the LUT inversion was then performed as described in sect. 2.3.1.

In this way, the speed of the LUT based inversion was comparable with the NN based approach. However, the resulting estimation accuracy decreased significantly (LAI:

RMSE=0.8, LCC: RMSE=17.4 μg cm^{-2} and CCC: RMSE=0.46 g m^{-2}) despite this high number of classes. This lower estimation accuracy in comparison to the pixel based approach may be due to the loss of spectral information, especially affecting high values of LAI. Enhancing the number of classes may increase the accuracy, but also the computation time.

Although the neural nets may be more suitable for such fast operations, a drawback relates to their reduced availability. For the current work, the networking was performed under the Matlab processing environment (The Mathworks, 2007). Unfortunately, standard image processing software does not yet include this mapping technique. To foster the use of physically based approaches, providers of image processing software should add modules for direct and inverse modelling.

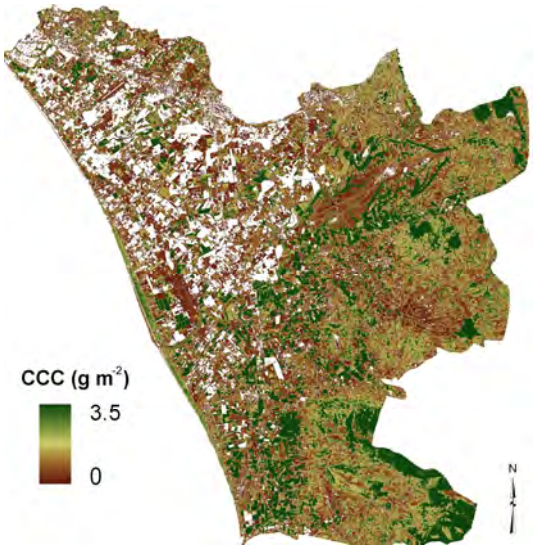


Figure 4. Spatial distribution of canopy chlorophyll content in the "Piana del Sele", Italian Campania region, derived from the neural network (white zones correspond to urban areas or greenhouses).

4. CONCLUSIONS

Recently available multispectral RapidEye data were tested for their operational suitability to estimate canopy biophysical variables in an agricultural area of Southern Italy. The physically based retrieval approaches outperformed the empirical methods, whereas the retrieval accuracy of the LUT was slightly better than the neural networks approach. However, the latter, already used in operational applications for coarse resolution data (Bacour et al., 2006), is much faster rendering it more suitable in this context. An unsupervised classification of the imagery prior to the RTM inversion was applied to reduce calculation time, as proposed by Baret and Buis (2008), but results were less accurate than pixel based procedures.

Generally, the canopy based variables (LAI and canopy chlorophyll content) could be estimated with much higher accuracy than variables on leaf level (leaf chlorophyll content), confirming previous findings of the literature (Baret et al., 2007). This might be a drawback for certain applications, where properties of the leaves are required. For general precision farming applications, however, information at the canopy level can be sufficient or even of advantage over the leaf level (Baret et al., 2007).

Conclusively, RapidEye sensor provides useful data to derive biophysical variables for operational applications in the agricultural sector. Such applications may include, for instance, the modelling of crop water requirements, needing LAI as input, or the assessment of plant nitrogen status, requiring the information of canopy chlorophyll content. The use of physically based approaches to estimate these variables is suggested. Further validation work is required to test the applicability of these techniques for different areas and crops.

REFERENCES

Asner, G.P., 1998. Biophysical and biochemical sources of variability in canopy reflectance. *Remote Sens. Environ.*, 64, pp. 234-253.

Atzberger, C., 2004. Object-based retrieval of biophysical canopy variables using artificial neural nets and radiative transfer models, *Remote Sens. Environ.*, 93, pp. 53-67.

Baret F., Houlès V. and Guérif, M., 2007. Quantification of plant stress using remote sensing observations and crop models: the case of nitrogen management, *J. Exp. Bot.*, 58(4), pp. 869 - 880.

Baret, F. and Buis, S., 2008. *Estimating canopy characteristics from remote sensing observations: Review of methods and associated problems*. In S. Liang (Ed.), Advances in Land Remote Sensing: System, Modeling, Inversion and Application, Springer, pp. 171−200.

Bacour, C., Baret, F., Béal, D., Weiss, M., & Pavageau, K., 2006. Neural network estimation of LAI, fAPAR, fCover and LAI × Cab, from top of canopy MERIS reflectance data: Principles and validation. *Remote Sens. Environ.*, 105, pp. 313−325.

Chen, J.M., Black, T.A., 1992. Defining leaf-area index for non-flat leaves. *Plant Cell. Environ.*, 15, pp. 421–429.

Clevers, J.G.P.W., 1989. The application of a weighted infrared-red vegetation index for estimating leaf area index by correcting for soil moisture. *Remote Sens. Environ*., 29, pp. 25–37.

Colombo, R., Bellingeri, D., Fasolini, D., & Marino, C. M. 2003. Retrieval of leaf area index in different vegetation types using high resolution satellite data. *Remote Sens. Environ.*, 86(1), pp. 120−131.

Combal, B., Baret, F., Weiss, M., Trubuil, A., Mace´, D., Pragne`re, A., et al., 2002. Retrieval of canopy biophysical variables from bidirectional reflectance using prior information to solve the ill-posed inverse problem. *Remote Sens. Environ.*, 84, pp. 1– 15.

Darvishzadeh, R., Skidmore, A., Schlerf, M., Atzberger, C., 2008. Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in a heterogeneous grassland. *Remote Sens. Environ.*, 112, pp. 2592-2604.

Dash, J., and Curran, P.J., 2007. Evaluation of the MERIS terrestrial chlorophyll index (MTCI). *Adv. Space Res.*, 39(1), pp. 100-104.

Dash, J., Curran, P. J., Tallis, M. J., Llewellyn, G. M., Taylor, G. and Snoeij, P., 2010. Validating the MERIS Terrestrial Chlorophyll Index (MTCI) with ground chlorophyll content data at MERIS spatial resolution. *International Journal of Remote Sensing*, in press.

De Michele, C., Vuolo, F., D'Urso, G., Marotta, L., Richter, K., 2009. The Irrigation Advisory Program of Campania Region: from research to operational support for the Water Directive in Agriculture. *Proc. of 33rd international Symposium on Remote Sens of Environ.*, May 4-8, Stresa, Italy.

D'Urso, G. and Calera, A., 2006. Operative Approaches to determine crop water requirements from earth observation data: methodologies and applications. *AIP Conference Proceedings, 852, pp. 14-25.*

Evans, J.R., 1989. Photosynthesis and nitrogen relationships in leaves of C3 plants. *Oecologia* 78, pp. 9-19.

Féret, J.-B., François, C., Asner, G.P., Gitelson, A.A., Martin, R.E., Bidel, L.P.R., Ustin, S.L., le Maire, G. and Jacquemond, S., 2008. PROSPECT-4 and 5: advances in the leaf optical properties model separating photosynthetic pigments. *Remote Sens. Environ.*, 112, pp. 3030–3040.

Glenn, E.P., Huete, A.R., Nagler, P.L., Nelson, S.G., 2008. Relationship between remotely-sensed vegetation indices, canopy attributes and plant physiological processes: What vegetation indices can and cannot tell us about the landscape. *Sensors*, 8, pp. 2136-2160.

Haboudane, D., Miller, J. R., Tremblay, N., Zarco-Tejada, Pablo J., Dextraze, L., 2001. Heterogeneity of CASI-estimated Chlorophyll Content: Assessment and Comparison with Ground Truth from L'ACADIE GEOIDE Experimental Site. *23rd Canadian Symposium on Remote Sensing*, Laval, Quebec (Canada), August 20th-24th, 2001. http://hdl.handle.net/10261/10653

Huemmrich, K. F., 2001. The GeoSail model: a simple addition to the SAIL model to describe discontinuous canopy reflectance. *Remote Sens. Environ.*,75 (3), pp. 423-431.

Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P.-J., Asner, G.P., François, C., Ustin, S.L., 2009. PROSPECT + SAIL models: A review of use for vegetation characterization. *Remote Sens. Environ.*, 113 (1), pp. 56-66.

Jonckheere, I., Fleck, S., Nackaerts, K., Muys, B., Coppin, P., Weiss, M. and Baret, F., 2004. Review of methods for in-situ leaf area index determination. Part I. Theories, sensors and hemispherical photography. *Agr. Forest Meteorol.*, 121, pp. 19–35.

Kuusk, A., 1991. *The hot spot effect in plant canopy reflectance*. In R. B. Myneni, & J. Ross (Eds.), Photon-vegetation interactions. Applications in optical remote sensing and plant ecology, Berlin: Springer Verlag, pp. 139−159.

Marchi, S., Sebastiani, L., Gucci, R., Tognetti, R., 2005. Sink–source Transition in Peach Leaves during Shoot Development. *J. Amer. Soc. Hort. Sci.* 130(6), pp. 928-935.

Migdall, S., Bach, H., Bobert, J., Wehrhan, M., Mauser, W., 2009. Inversion of a canopy reflectance model using hyperspectral imagery for monitoring wheat growth and estimating yield. *Precision Agric.*, 10 (6), pp. 508-524.

Moran, M.S., Mass, S.J., Pinter Jr., P.J., 1995. Combining remote sensing and modeling for estimating surface evaporation and biomass production. *Remote Sens. Rev.*, 12, pp. 335–353.

Neumann, H.H., Den Hartog, G.D., Shaw, R.H., 1989. Leaf-area measurements based on hemispheric photographs and leaf-litter collection in a deciduous forest during autumn leaf-fall. *Agric. For. Meteorol.*, 45, pp. 325–345.

Richter, K., Atzberger, C., Vuolo, F., D'Urso, G., 2009. Experimental assessment of the Sentinel-2 band setting for RTM-based LAI retrieval of sugar beet and maize. *Can. J. Rem. Sens.*, 35 (3), pp. 230-247.

Richter, R., 1998. Correction of satellite imagery over mountainous terrain. *Appl. Optics*, 37, pp. 4004-4015.

SPARC data acquisition report, 2004. Contract no: 18307/04/NL/FF, University Valencia.

Verhoef, W., 1984. Light scattering by leaf layers with application to canopy reflectance modeling: the SAIL model. *Remote Sens. Environ.*, 16, pp. 125−141.

Verhoef, W., 1985. Earth observation modeling based on layer scattering matrices. *Remote Sens. Environ.*, 17, pp. 165-178.

Weiss, M., Baret, F., Myneni, R.B., Pragnère, A., and Knyazikhin, Y., 2000. Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data. *Agronomie*, 20, pp. 3-22.

ACKNOWLEDGEMENTS

Funding for this research was provided by ESA under the MTCI-EVAL project, additional information available at: http://www.sen2chl.co.uk. Logistical support from Ariespace srl spin-off company (http://www.ariespace.com) was crucial in the success of the field campaign. Satellite data were funded by the University of Naples "*Federico II*" and acquired in the context of Irrigation Advisory Services of Campania Region (http://www.consulenzairrigua.com).

PANSHARPENING – RELATING QUANTITATIVE QUALITY MEASURES TO IMPACT ON RESULTS OF SUBSEQUENT PROCESSING STEPS

U. Weidner

Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology (KIT),
Englerstraße 7, 76128 Karlsruhe, Germany – uwe.weidner@kit.edu

KEY WORDS: Classification, Extraction, Fusion, Feature, Accuracy

ABSTRACT:

Since the advent of high-resolution satellite sensors the pansharpening of multispectral data using the higher spatial information of the panchromatic channel became a prominent topic in the data fusion community. Besides the development of new algorithms also the evaluation of approaches has been addressed in recent years leading e.g. to competitions as by IGARSS (Alparone et al., 2007) using visual inspection and quantitative measures for quality assessment. (Zhang, 2008) questioned the significance of such quantitative measures to evaluate the impact of pansharpening on subsequent processing like classification using an example based on linearly transformed data. This was the motivation for our investigations addressing the question, if quantitative evaluation criteria for pansharpening can reflect its impact on subsequent processing.

1. INTRODUCTION

The question addressed in this contribution is, if quantitative quality measures for pansharpening evaluation really indicate the quality of the data as input for subsequent processing like classification and its impact on the quality of the results of this processing. Fusion of multispectral data with data of higher geometric resolution has been already addressed for a long time. Nonetheless with the advent of high resolution satellite sensors like IKONOS and QuickBird a variety of different approaches based also on quite different techniques (c.f. (Zhang, 2004)) has been proposed taking also sensor characteristics into account in order to improve the results of pansharpening and thereby facing tightened requirements on the spectral consistency of the pansharpened multispectral data. As a consequence of both developments quality assessment has also been addressed since some years which led to the proposal of a number of quality measures and competitions of pansharpening approaches as e.g. by IGARSS (c.f. (Alparone et al., 2007)) combining visual / qualitative and quantitative evaluation. First the quantitative measures based on a comparison of the original and pansharpened multispectral data took mainly the spectral consistency into account. Recent developments consider not only the spectral consistency, but also the information content transferred from the panchromatic to the pansharpened multispectral data (c.f. (Alparone et al., 2006)). (Zhang, 2008) questioned the significance of such quantitative measures – focussing on the spectral consistency – with respect to the impact of the pansharpening on the results of subsequent processing using as example linearly transformed data and clustering: Although the clustering leads to the same results, the quantity Q4 (Alparone et al., 2004) used for quality assessment differs. In this case these differences for Q4 are due to different signal means and variances caused by the linear transformations, whereas the clustering is not sensitive and thereby the results are not influenced by these transformations. Nonetheless other subsequent processing and classification algorithms may be sensitive to distortions caused by the pansharpening.

The discussion above is the motivation of our recent investigations. We used simple pansharpening approaches – simple with respect to their concept and their implementation within toolboxes of image processing and remote sensing software packages – shortly outlined in the following section. Quantitative measures to evaluate the quality of the obtained pansharpened data are presented and discussed in Section 3. The pansharpened data sets are then used for clustering and feature extraction. Based on the results, the impact of pansharpening on clustering and feature extraction is shown and evaluated in Section 4. followed by conclusions and an outlook.

2. PANSHARPENING APPROACHES

Simple pansharpening approaches like Brovey or principal component analysis (PCA)-based approaches suffer from the fact that the following two conditions

$$C_{pan} \propto \sum_{j}^{n} w_j C_j \tag{1}$$

with C_{pan} denoting the panchromatic channel, C_j a multispectral channel and w_j its weight and

$$C_{pan} \propto C_{sub} \tag{2}$$

with C_{sub} the channel to be substituted and high correlation between the two channels is required, but not met. The first condition (1) is valid for sensor systems like QuickBird and is of importance for arithmetic combinations like Brovey on one hand and frequency-based algorithms on the other. This condition is e.g. used by used by (Kalpoma and Kudoh, 2007) and (Aiazzi et al., 2007). From (1) also follows the existence of a linear transform to map the panchromatic channel to the weighed sum. An example for such a linear transformation is histogram matching based on the means and the standard deviations of the signals. The second condition (2) is required for approaches based on component substitution by the high-resolution channel: the component to be substituted and the panchromatic channel have to be similar.

In (Weidner and Centeno, 2009) we investigated four approaches with respect to their capability to maintain spectral consistency. The first approach is an adapted Brovey transformation (BROV) taking into account the different spectral bandwidths of the channels to compute a weighed intensity channel as a lower resolution approximation of the high resolution panchromatic channel and performing a linear transformation based on histogram matching between these before the arithmetic combination. The second

approach investigated is a PCA-based approach based on substitution. This approach was mainly included as a kind of reference in order to show the limitations of PCA-based approaches and to derive an approach based on an orthogonal transformation. The idea of this third approach (ORTHO) is to define a linear combination of the multispectral channels according to the conditions stated above for the channel to be substituted, thus defining the first row of the transformation matrix. The other rows are constructed in such a way that the rows constitute an orthogonal basis. The fourth approach is based on linear filter, namely Laplace and is thereby frequency-based (PANSHLPL). Details on the implementations of these approaches can be found in (Weidner and Centeno, 2009).

3. QUANTITATIVE EVALUATION APPROACHES

(Wang and Bovik, 2002) proposed the index defined by

$$\rho_{WB} = \frac{4 \cdot \sigma_{AB} \cdot \mu_A \cdot \mu_B}{(\sigma_A^2 + \sigma_B^2)(\mu_A^2 + \mu_B^2)} \tag{3}$$

ranging from 0 to 1 to measure the similarity between two one-channel gray-valued images. μ_A and μ_B denote the means, σ_A^2 and σ_B^2 the variances and σ_{AB} the covariance of the images. It can be rewritten to

$$\rho_{WB} = \frac{\sigma_{AB}}{\sigma_A \cdot \sigma_B} \cdot \frac{2 \cdot \mu_A \cdot \mu_B}{\mu_A^2 + \mu_B^2} \cdot \frac{2 \cdot \sigma_A \cdot \sigma_B}{\sigma_A^2 + \sigma_B^2} \tag{4}$$

in order to clarify the parameters. Each criterion has been used in a number of publications on pansharpening evaluation, but they are now combined. A detailed discussion of terms and how (4) can also be used within pansharpening is given in (Weidner and Centeno, 2009). In a first step (Alparone et al., 2004) generalised the index of (Wang and Bovik, 2002). This generalised index Q4 is applied in (Alparone et al., 2007) besides two other measures - SAM and ERGAS. Their generalisation is based on the use of quaternions and thereby restricted to evaluate results of images with $n = 4$ channels. We therefore propose to generalise the index of (Wang and Bovik, 2002) by

$$\rho^*_{WB} = \frac{4 \cdot \mathrm{tr}(\Sigma_{AB}) \cdot |\underline{\mu}_A| \cdot |\underline{\mu}_B|}{(\mathrm{tr}(\Sigma_A) + \mathrm{tr}(\Sigma_B))(|\underline{\mu}_A|^2 + |\underline{\mu}_B|^2)} \tag{5}$$

where tr denotes the trace of a matrix, $\underline{\mu}_A$ the vector of mean values of data set A, $|\underline{\mu}_A|$ the length of the vector, Σ_A the covariance matrix of A, and Σ_{AB} the covariance matrix of the data sets to be compared. Instead of the quantity $Q4$, (5) is not restricted to four channels. In order to overcome the restriction of dimension and to also include the information content of the pansharpened multispectral channels in the quality assessment, (Alparone et al., 2006) proposed a new index QNR which consists of two terms. The first term is based on the quality index of (Wang and Bovik, 2002) comparing the multispectral original and pansharpened data and added for all channels, the second is based on similarity measure between multispectral and panchromatic data in a high resolution and low resolution version. In our investigation we will use a different approach motivated by (Xydeas and Petrovic, 2000) to measure the information content of the pansharpened image, namely the correlation of gradient information. For this purpose first a weighed mean of the multispectral data according to (1) is computed followed by computing the gradient absolute values of the weighed mean of the multispectral channels and the panchromatic channel and their correlation ρ^*_∇ leading to the quantity

$$q^*_T = q^*_{WB} \cdot \rho^*_\nabla \tag{6}$$

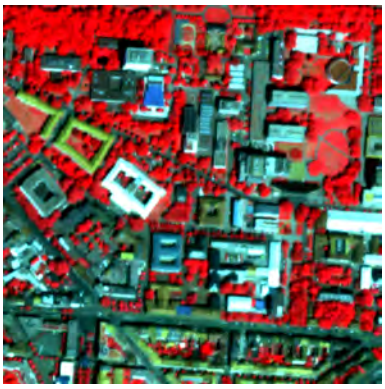


Figure 1: Subset URBAN - original data

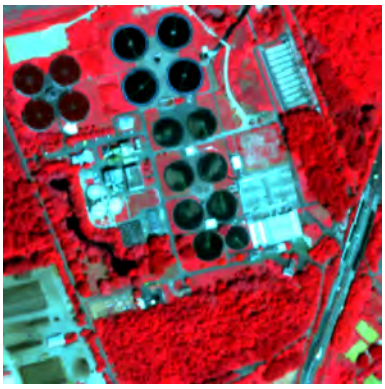


Figure 2: Subset FOREST - original data

Figure 3: Subset RIVER - original data

as a combined quantity to evaluate spectral consistency and information content. We prefer this quantity, because the gradient information forms the basis for point / feature extraction and should therefore be closely related to results of these algorithms.

4. SETUP OF INVESTIGATIONS AND RESULTS

Our investigations relating quantitative quality measures for pansharpening with the impact of pansharpening on subsequent processing of remote sensing data is based on a number of subsets from a QuickBird scene. Results for three subsets are discussed in more detail. These subsets constitute of an urban area (URBAN, Fig. 1), a mixed area with a sewage plant and forest (FOREST, Fig. 2) and an area with a larger river bed (RIVER, Fig. 3). The first subset provides a variety of different surface covers including man-made structures. The other subsets are selected for their simplicity on one hand, but also for their spectral properties – namely the homogeneous forest and water areas – on the other hand.

URBAN	Layer	q_{WB}, q^*_{WB}	ρ^*_C	q^*_T
BROV	1	0,877	0,920	0,807
	2	0,918	0,931	0,855
	3	0,953	0,896	0,854
	4	0,952	0,906	0,862
	Mean	0,925	0,913	0,844
	Matrix	0,939		0,857
	SAM	0,469		
PCA	1	0,691	0,995	0,687
	2	0,700	0,998	0,698
	3	0,703	0,994	0,699
	4	0,990	0,572	0,567
	Mean	0,771	0,890	0,663
	Matrix	0,855		0,761
	SAM	5,167		
ORTHO	1	0,900	0,936	0,842
	2	0,966	0,906	0,875
	3	0,960	0,916	0,880
	4	0,934	0,953	0,891
	Mean	0,940	0,928	0,872
	Matrix	0,944		0,876
	SAM	1,468		
PANSHLPL	1	0,933	0,883	0,824
	2	0,956	0,886	0,847
	3	0,975	0,850	0,828
	4	0,976	0,835	0,815
	Mean	0,960	0,863	0,829
	Matrix	0,968		0,836
	SAM	0,000		
FOREST	Layer	q_{WB}, q^*_{WB}	ρ^*_C	q^*_T
BROV	1	0,872	0,892	0,777
	2	0,920	0,893	0,821
	3	0,963	0,827	0,796
	4	0,966	0,907	0,876
	Mean	0,930	0,879	0,818
	Matrix	0,957		0,841
	SAM	0,677		
PCA	1	0,377	0,996	0,376
	2	0,426	0,997	0,424
	3	0,385	0,996	0,384
	4	0,954	0,418	0,399
	Mean	0,536	0,852	0,396
	Matrix	0,759		0,647
	SAM	10,979		
ORTHO	1	0,878	0,921	0,809
	2	0,958	0,899	0,861
	3	0,959	0,893	0,856
	4	0,964	0,951	0,917
	Mean	0,940	0,916	0,860
	Matrix	0,959		0,879
	SAM	1,165		
PANSHLPL	1	0,921	0,862	0,794
	2	0,953	0,858	0,818
	3	0,978	0,793	0,776
	4	0,984	0,836	0,823
	Mean	0,959	0,837	0,803
	Matrix	0,977		0,818
	SAM	0,000		
RIVER	Layer	q_{WB}, q^*_{WB}	ρ^*_C	q^*_T
BROV	1	0,847	0,885	0,749
	2	0,903	0,892	0,805
	3	0,963	0,798	0,768
	4	0,989	0,958	0,947
	Mean	0,925	0,883	0,817
	Matrix	0,983		0,868
	SAM	0,457		
PCA	1	0,441	0,972	0,429
	2	0,483	0,990	0,478
	3	0,576	0,968	0,557
	4	0,969	0,783	0,759
	Mean	0,617	0,928	0,556
	Matrix	0,938		0,871
	SAM	3,938		
ORTHO	1	0,856	0,916	0,784
	2	0,945	0,912	0,862
	3	0,939	0,904	0,849
	4	0,989	0,970	0,959
	Mean	0,932	0,925	0,863
	Matrix	0,984		0,911
	SAM	1,061		
PANSHLPL	1	0,939	0,831	0,781
	2	0,963	0,840	0,809
	3	0,985	0,748	0,738
	4	0,996	0,901	0,897
	Mean	0,971	0,830	0,806
	Matrix	0,993		0,825
	SAM	0,000		

Table 4: Evaluation by quantitative measure

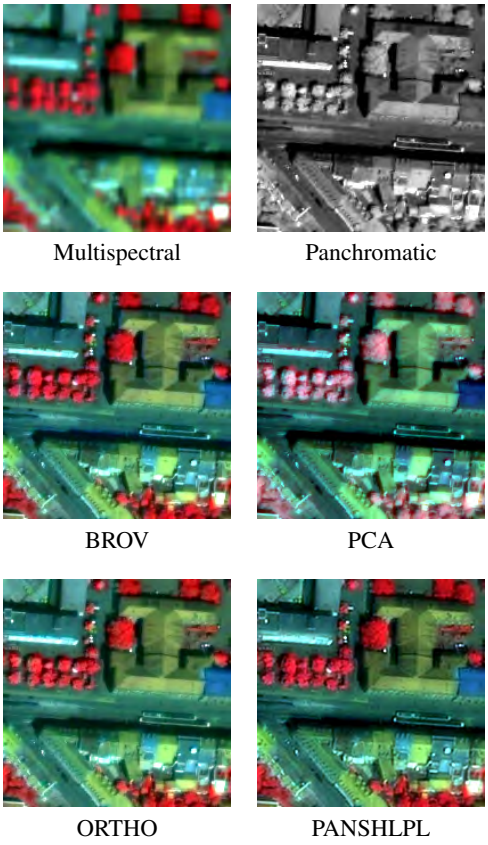


Figure 5: Results for subset URBAN (section)

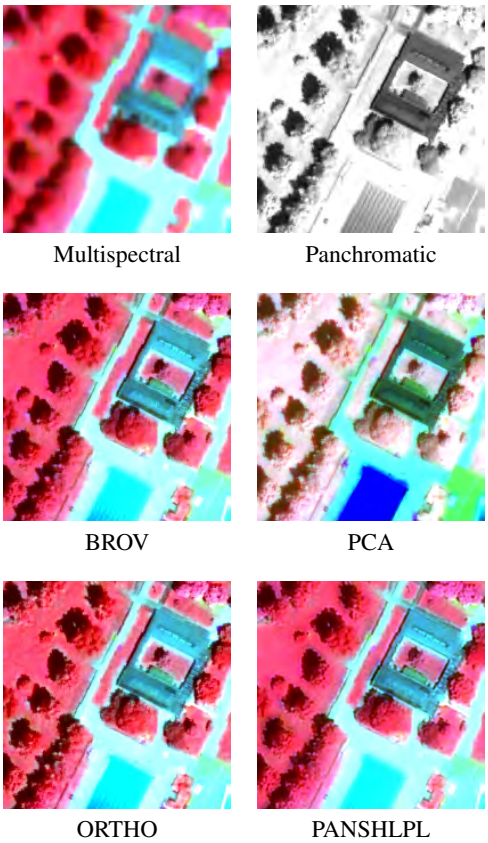


Figure 6: Results for subset RIVER (section)

URBAN	Entire			
	Overall	Kappa		
BROV	0,741	0,714		
PCA	0,750	0,723		
ORTHO	0,754	0,727		
PANSHLPL	0,823	0,800		
	Homogeneous		Non-homogeneous / edges	
	Overall	Kappa	Overall	Kappa
BROV	0,830	0,804	0,668	0,642
PCA	0,824	0,797	0,670	0,644
ORTHO	0,837	0,811	0,663	0,636
PANSHLPL	0,909	0,892	0,729	0,704

Table 7: Comparison of clustering results URBAN

Tab. 4 compiles the quantitative indicators discussed in the previous section in sequence of the three subsets and the algorithms described in Section 2.. The results are given for the four single multispectral layers (B, G, R, NIR) numbered from 1 to 4 according to (3), the mean value of these results and the combination of the four channels using (5). The next column compiles the correlation of corresponding Soobel gradient images and the last column the results for q_T^* according to (6). In this contribution the quantities are computed for the entire subsets. A distinction between homogeneous and non-homogeneous areas (edges) is made in (Weidner and Centeno, 2009). The ranking of algorithms keeps the same, but results are worse within edge regions.

Fig. 5 displays the original data and the results for a smaller section – using a clipping of 5 % of each channels histogram – of the subset URBAN. The quantities involved in (6) show only minor differences except PCA. For all algorithms the result for the first channel (B) is minor to the other channels. A closer analysis further indicates that those algorithms mainting spectral properties are minor in transferring the gradient information to the pansharpened multispectral channels. This trade-off is not surprising especially for simple algorithms as used here. Furthermore, those algorithms performing a linear transformation work slightly better than others taking the correlation of gradient information into account. Nonetheless, the quantities indicate problems with the PCA-based algorithm, which also could be verified by visual inspection. In fact the PCA-based algorithm turns out to perform worst (c.f. Fig. 6) also for the two other selected subsets due to their special characteristics and the data dependence of PCA determining the transformation matrix. The difficulties for single channels are disguised by the criterion based on the set of all channels. The SAM criterion indicates that the fourth algorithm based on linear filtering has no impact on SAM independent from the data.

In the following we address the impact of pansharpening on two possible processing steps, namely classification and point extraction. We selected these two processing steps, because classification depends on the spectral properties and point extraction on the information content given by the gradient information.

For unsupervised classification we selected *k-means* clustering and 8 clusters for all subsets although the complexity of the subsets is different. The clustering results for the different pansharpening results are not identical concerning their ID and therefore clusters are automatically assigned using the overall accuracy as criterion. Clustering adopts for linear changes in feature space and therefore linear spectral transformations do not have an influence on its result. For the comparison of the clustering results using the pansharpened data of the first data set URBAN Tab. 7 compiles the overall accuracies and the κ indices. The clustering is performed on the entire data set, but for the analysis three cases a discerned: entire data set, homogeneous and non-homogeneous areas as in (Weidner and Centeno, 2009) for the quality measures applied therein. Within the homogeneous areas the results

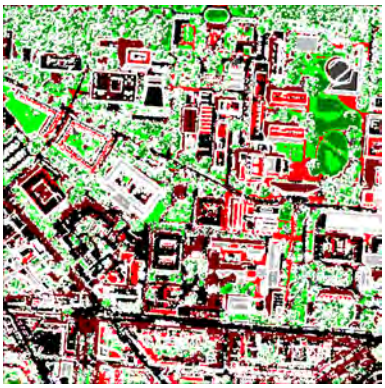


Figure 8: Subset URBAN clustering - BROV homogeneous areas

	FOREST		RIVER	
	Overall	Kappa	Overall	Kappa
BROV	0,723	0,690	0,911	0,852
PCA	0,660	0,627	0,934	0,891
ORTHO	0,717	0,684	0,868	0,780
PANSHLPL	0,842	0,821	0,948	0,913

Table 9: Comparison of clustering results FOREST / RIVER (homogeneous areas)

of clustering should not differ, for edge regions differences have to be expected. As an example the result of clustering based on the pansharpened data by BROV within the homogeneous areas is shown in Fig. 8. Although the overall visual impression for the clustering results is almost the same, smaller deviations can be observed which also lead to a lower overall accuracy taking the clustering result of the original multispectral data as reference. The ranking is the same for all three cases mentioned above taking the significance of differences between the overall accuracies into account. Again the result for the edge regions are minor to those of the homogeneous regions.

The results for the first data set are confirmed by the results of the subset FOREST (c.f. Tab. 9). In case of the subset RIVER however the ranking according the overall accuracy differs significantly. The visual comparison of the data (c.f. Fig. 6 and Fig. 10 for the results of clustering) is in accordance with the quality evaluation taking the single channels as basis, but is not apparent neither by ρ_{WB}^* nor the overall accuracy. The value and / or range of overall accuracies depend on the data and the class distribution. The data set was selected to study homogeneous regions (dominant cluster river) and their impact on the preprocessing for pansharpening. Therefore not the absolute values but only the rankings are of interest.

For the subset RIVER a test of the impact of pansharpening on Maximum-Likelihood (ML) classification is accomplished. Compared to the other data sets it comprises only few classes. Four training areas for the river (Rhine), fields (bare soil), meadows and forest – the most dominant classes – are defined in homogeneous areas. The first set of classifications used the spectral signatures from the original multispectral data serving as spectral library, the second set the spectral signatures from each data set. Therefore the results of the first set are affected by linear changes, but not those of the second set. All other parameter settings are kept the same. Results are displayed in Fig. 11 showing only a small section for better visualisation and statistics for the entire subsets (homogeneous areas) are given Tab. 12. Except for PCA the achieved overall accuracies are in the same range for both settings. The deviations for PCA are caused by the spectral distortions (c.f. Tab. 4) and are anticipated (c.f. Tab. 13 for the

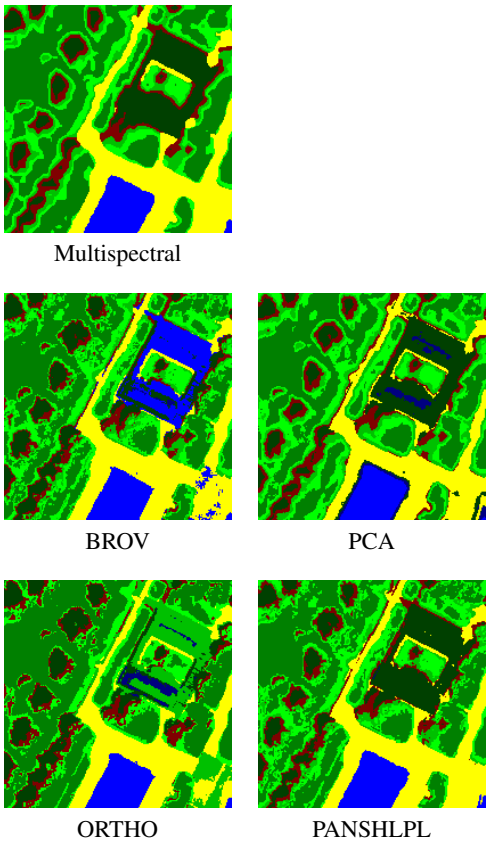


Figure 10: Clustering results for subset RIVER (section)

Bhattacharrya distances between the original spectral data and those of the pansharpened for each training area).

In order to evaluate the impact of pansharpening algorithms on point extraction the SIFT operator (Lowe, 2004) is selected including also a matching of points across the data sets. The evaluation is based on the pansharpened multispectral data for which a weighed mean according to (1) is computed. Point extraction uses the panchromatic channel as reference and only those points which are matched across all pansharpening results for a data set are included in the analysis. The statistics in Tab. 14 are based on about 260 points for the subset URBAN, 280 points for the subset FOREST and 140 points for the subset RIVER. These points are evenly distributed across the data sets except for the river which also caused the lower number of matched points. Given are the

Spectral signature from original MS	Overall	Kappa
BROV	0,951	0,905
PCA	0,221	0,208
ORTHO	0,947	0,898
PANSHLPL	0,975	0,951

Spectral signature from each data set	Overall	Kappa
BROV	0,949	0,901
PCA	0,835	0,693
ORTHO	0,921	0,849
PANSHLPL	0,957	0,917

Table 12: Comparison ML-classification (homogeneous areas)

	Rhine	fields	meadows	forest
BROV	0,89	1,60	0,66	0,54
PCA	44,95	25,55	30,89	3,25
ORTHO	0,89	1,72	0,52	0,65
PANSHLPL	0,24	0,05	0,07	0,46

Table 13: Bhattacharrya distances

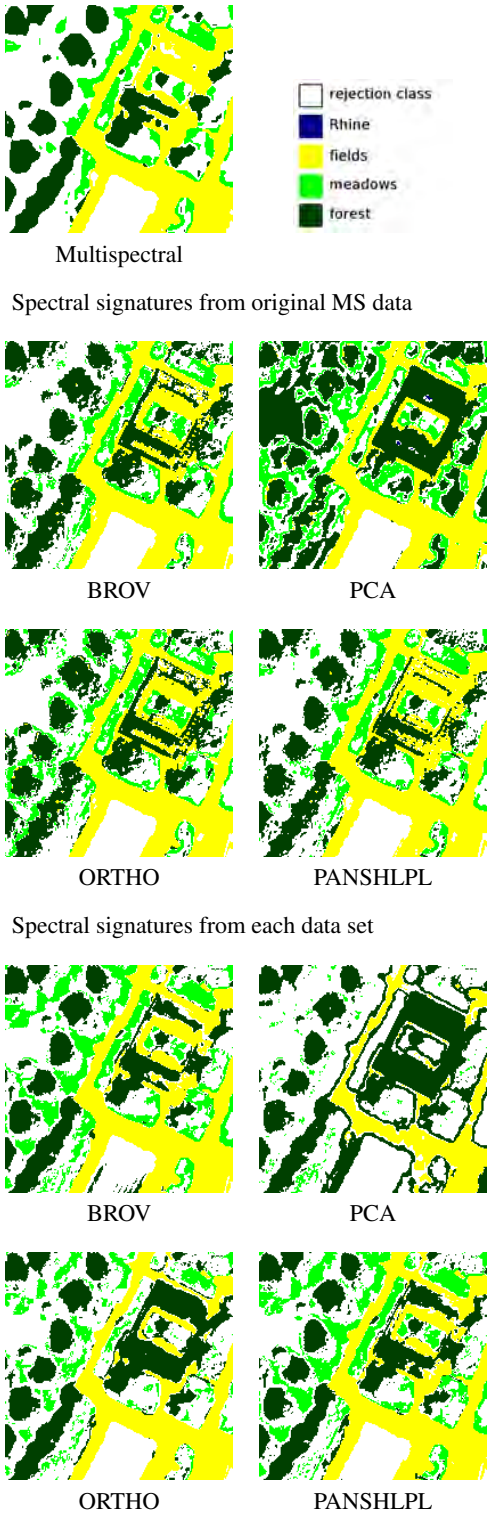


Figure 11: ML-classification results for subset RIVER (section)

URBAN	mean ds	sigma ds	median ds
MS (original data)	2,1	34,4	1,9
BROV	0,2	2,5	0,1
PCA	1,4	22,2	1,1
ORTHO	0,1	2,0	0,0
PANSHLPL	1,7	28,3	1,6

FOREST	mean ds	sigma ds	median ds
MS (original data)	1,9	32,0	1,8
BROV	0,3	4,9	0,1
PCA	1,7	28,3	1,4
ORTHO	0,3	5,1	0,1
PANSHLPL	1,6	27,2	1,5

RIVER	mean ds	sigma ds	median ds
MS (original data)	2,3	26,7	1,9
BROV	0,2	1,8	0,1
PCA	2,0	23,7	1,4
ORTHO	0,2	2,2	0,1
PANSHLPL	2,1	24,3	1,8

Table 14: Results of point matching (SIFT) $[pixel]$

mean point distances, their standard deviation and their median for comparison. A low mean and median of point distances indicate that the matched point of the pansharpened data are close to the original positions derived from the high resolution panchromatic channel. If furthermore the standard deviation of the point distances is low, most point positions are in agreement. The correlations of the Sobel gradients given in Tab. 4 are almost in the same range for all algorithms and data sets. For comparison the original multispectral data is also included in Tab. 14. The correlations for these unsharpened data sets are about 0.4 and therefore the high values for the point deviations for the unsharpened data sets and those of BROV and ORTHO are in agreement, but not those of PCA and PANSHLPL. For the PANSHLPL result the reason may be found in effects due to the unsharp masking leading to visual sharp, but deteriorated edge information. The PCA and ORTHO algorithms are similar in the sense that they both perform a linear transformation, nevertheless the spectral distortions produced by the PCA algorithms cause problems within matching and finally in the point positions. These results with respect to its ranking are confirmed by recent preliminary results of point extraction based on manual measurement and least squares matching for point transfer.

5. CONCLUSIONS

Pansharpening and its evaluation using quantitative measures are addressed by a large number of publications. Motivated by (Zhang, 2008) this contribution investigates to which extend such quality measures reflect the impact of pansharpening on subsequent processing steps in remote sensing. For this purpose simple pansharpening approaches are applied to QuickBird data. As examples for processing classification and point extraction are investigated. The quality measures used here evaluate mostly linear spectral distortion and the degree of information transformation from the high resolution panchromatic channel to the multispectral channels, namely the gradient. Partly subsequent processing steps are insensitive to linear transformations like clustering at least within homogeneous regions. Nevertheless, problems occur at edges. Therefore homogeneous and non-homogeneous regions should be evaluated separately (c.f. (Perko, 2004) and (Weidner and Centeno, 2009)). In most cases it could be demonstrated that the quality indices for pansharpening evaluation are in agreement with the quality of the processing results based on the pansharpened data, but also cases are discussed where the quality measures for the pansharpening results do not significantly differ, but the results of the processing. The reason for this is the fact that processing steps are influenced by effects of pansharpening which can not entirely evaluated by a single quantity – although a combination of different criteria. Therefore, just one index seems to be inappropriate. In addition specialised criteria with respect to input data and processing steps should also be used like SAM for hyperspectral data (Klonus and Ehlers, 2009). The results of our investigations are based on a number of subsets. Therefore tests will be performed on further data sets in order to form a wider basis. Furthermore the impact on segmentation will be addressed based on (Weidner, 2008).

REFERENCES

Aiazzi, B., Baronti, S. and Selva, M., 2007. Improving component substitution pansharpening through multivariate regression of MS+pan data. IEEE Transactions on Geoscience and Remote Sensing 45(10), pp. 3230 – 3239.

Alparone, L., Aiazzi, B., Baronti, S., Garzelli, A. and Nencini, F., 2006. Information-theoretic image fusion assessment without reference. In: ESA-EUSC 2006: Image Information Mining for Security and Intelligence.

Alparone, L., Baronti, S., Garzelli, A. and Nencini, F., 2004. A global quality measurement of pan-sharpened multispectral imagery. IEEE Geoscience and Remote Sensing Letters 1(4), pp. 313 – 317.

Alparone, L., Wald, L., Chanussot, J., Thomas, C., Gamba, P. and Bruce, L., 2007. Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data-fusion contest. IEEE Transactions on Geoscience and Remote Sensing 45(10), pp. 3012 – 3021.

Kalpoma, K. and Kudoh, J.-I., 2007. Image fusion processing for IKONOS 1-m color imagery. IEEE Transactions on Geoscience and Remote Sensing 45(10), pp. 3075 – 3086.

Klonus, S. and Ehlers, M., 2009. Auswirkungen der Bildfusion auf Hyperspektraldaten. In: 29. Wissenschaftlich-Technische Jahrestagung der DGPF 2009, Jena, pp. 277 – 286.

Lowe, D., 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60, pp. 91 – 110.

Perko, R., 2004. Computer Vision for Large Format Digital Aerial Cameras. PhD thesis, Institute for Computer Graphics and Vision, Graz University of Technology.

Wang, Z. and Bovik, A., 2002. A universal image quality index. IEEE Signal Processing Letters 9(3), pp. 81 – 84.

Weidner, U., 2008. Contribution to the assessment of segmentation quality for remote sensing applications. In: IAPRSIS, Vol. 37, Part B7, pp. 1101 – 1110.

Weidner, U. and Centeno, J., 2009. Pansharpening – simple approaches and their evaluation. Photogrammetrie - Fernerkundung - Geoinformation PFG 4, pp. 317 – 327.

Xydeas, C. and Petrovic, V., 2000. Objective image fusion performance measure. Electronic Letters 36(4), pp. 308 – 309.

Zhang, Y., 2004. Understanding image fusion. Photogrammetric Engineering & Remote Sensing 70(6), pp. 657 – 661.

Zhang, Y., 2008. Methods for image fusion quality assessment - a review, comparison and analysis. In: IAPRSIS, Vol. 37, Part B7, pp. 1101 – 1110.

MAPPING DEFORMATION OF MAN-MADE LINEAR FEATURES USING DINSAR TECHNIQUE

H. Wu [a, *], Y. Zhang [a], J. Zhang [a], X. Chen [b]

[a] Key Laboratory of Mapping from Space of State Bureau of Surveying and Mapping, Chinese Academy of Surveying and Mapping, Beijing 100830, China – wha_105@yahoo.com.cn, (yhzhang, zhangjx)@casm.ac.cn
[b] East China Institute of Technology, Nanchang 330013, China - chenxy@ecit.edu.cn

KEY WORDS: Monitoring, SAR, Technology, Image, Radar

ABSTRACT:

Man-made linear features like dams, highways, airports, and so on are very important infrastructures in any society. However, due to natural and human activities, ground deformation is threatening many linear features all over the world. Because of groundwater over-exploration, land subsidence has taken place in Taiyuan, China for many years. The South Ring Expressway and Wushu International Airport in the southern suburb of Taiyuan, is being threaten by severe ground subsidence. In this paper, 13 Envisat Advanced Synthetic Aperture Radar (ASAR) images from 2006 to 2009 have been acquired to monitor the southern Taiyuan City using small baseline differential interferometric SAR (DInSAR) technique. Along the South Ring Expressway, the maximum deformation lies in the central section, with an average velocity about -50 ~ -90 mm/a. The eastern section has the minimum deformation about -15 mm/a while the western section has a velocity about -10 ~ -50 mm/a. Over the Wushu International Airport, the deformation velocity is around -42 mm/a in the southwest and -20 mm/a in the northeast. Precise levelling data in the study area validate our results.

1. INTRODUCTION

Man-made linear features include dams, highways, railways, airports, oil or gas pipelines, and others. Most of them are vital infrastructures and facilities to sustain normal operation of any society. However, affected by natural and human activities, deformation is threatening many linear features all over the world. The deformation not only weakens the function of these man-made features, but also threatens the human life. Therefore, it is necessary to monitor the deformation timely and accurately and take corresponding measures to control it. Traditional ways of monitoring ground deformation utilize levelling and global positioning system (GPS) measurement. They have the advantage of high accuracy. However, they are very costly and can not achieve enough spatial sampling density.

With the development of radar remote sensing technique, differential interferometric SAR (DInSAR) is playing an important role in monitoring ground deformation. However, for long term deformation, conventional 2-pass or 3-pass DInSAR can be easily affected by temporal or geometric de-correlation which degrades the interferometric phase and makes it difficult to extract useful information (Ferretti, 1999). To solve this problem, some advances in this field have been introduced based on point target analysis, including permanent scatterers (PS) method (Ferretti, 2000, 2001; Vilardo, 2009), small baseline subsets (SBAS) method (Berardino, 2002, 2003a, 2003b, 2004; Casu, 2008) and coherent points (CP) method (Mora, 2003).

In this paper, the deformation in Taiyuan is investigated with time-series DInSAR analysis. Taiyuan, the capital of Shanxi Province, China, has been a city suffering water shortage severely. Over exploring groundwater has taken place in this city over 50 years, resulting in 4 significant subsidence centres in the urban area, namely Xizhang, Xiayuan, Wanbolin and Wujiapu (Sun, 2007). In recent years, with fast development of economy in the southern suburb, particularly the exploitation of Taiyuan Economic and Technological Development Zone, new subsidence centre has been found in this area, where the South Ring Expressway and Wushu International Airport are located. Therefore, it is urgent to monitor the deformation of this area. In this research, deformation maps around the South Ring Expressway and Wushu International Airport in Taiyuan are generated for the first time with small baseline DInSAR technique by combining PS and CP methods.

2. STUDY AREA AND DATASETS

2.1 Study area

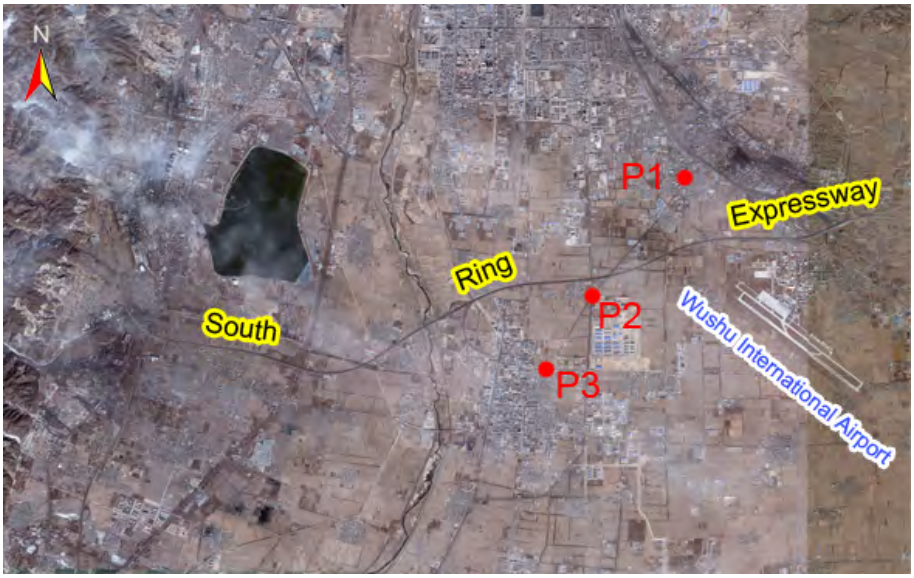
The study area, Taiyuan, is located in the middle reaches of the Yellow River in North China. The Fenhe River, a key tributary of the Yellow River, runs through the territory of Taiyuan City over a distance of about 100 km. Taiyuan lies in a basin bounded by the Taihang Mountain Ranges in east and the Lvliang Mountain Ranges in west. The annual average precipitation in Taiyuan is 479 mm. Taiyuan is one of the cities lacking of water resources in China. The annual average capacity of water resource is 5.7×10^8 m^3 in this city. And the annual amount of water resources per capita is 172 m^3, which accounts for 46% of that in the whole province, or 6.6% of that in China. This is also the reason of over-exploring ground water in Taiyuan for a long time.

* Corresponding author.

2.2 Datasets

13 ENVISAT advanced synthetic aperture radar (ASAR) single look complex (SLC) images over the study area are acquired from February 2006 to February 2009. The temporal baselines and normal baselines are shown in Table 2. To remove the topographic phase from interferograms, digital elevation model (DEM) generated by Shuttle Radar Topography Mission (SRTM) with 3 arc-second resolution (about 90m) is used. In addition, 3 precise levelling measurements (Figure 1(b)) are utilized to validate the DInSAR experimental results.

(a)



(b)

Figure 1. Study area: (a) location, (b) optical image provided by Google Earth with levelling points (red)

3. METHOD

There are two parts contained in the small baseline DInSAR technique, including linear deformation retrieval and non-linear deformation retrieval. Let us start our analysis by considering N SAR images acquired at the ordered times $(t_1,\cdots,t_N)$. We assume that each acquisition may interfere with at least another image and the normal baselines are small enough (e.g., 400m is chosen in this study) to reduce geometrical de-correlation. Based on N SAR images, we can generate M interferograms.

After removing the flat phase and topographic phase, each differential phase can be obtained, which can be expressed as

$$\phi_{dif\_i} = \phi_{def\_i} + \phi_{errortopo\_i} + \phi_{atm\_i} + \phi_{noise\_i} \quad (1)$$

where i is the i th interferogram, ϕ_{def} is the component due to the displacement of the terrain in satellite's look direction

between two SAR acquisitions, including linear deformation and non-linear deformation; $\phi_{errortopo}$ is the phase component

#	Date	Temporal Baseline /day	Normal baseline /m
1	2006-2-12	0	0
2	2006-11-19	280	-175
3	2007-1-28	350	-55
4	2007-3-4	385	759
5	2007-5-13	455	275
6	2007-9-30	595	223
7	2007-11-4	630	643
8	2007-12-9	665	-52
9	2008-1-13	700	403
10	2008-2-17	735	-2
11	2008-8-10	910	675
12	2008-12-28	1050	395
13	2009-2-1	1085	128

Table 2. ASAR images used in this research

due to the height error; ϕ_{atm} is the phase related with atmospheric artefacts; ϕ_{noise} is the noise phase including thermal noise, temporal and spatial decorrelation. The terms ϕ_{def} and $\phi_{errortopo}$ can be written as follows:

$$\phi_{errortopo} = \frac{4\pi}{\lambda} \cdot \frac{b}{r \cdot \sin\theta} \cdot \varepsilon \tag{2}$$

$$\phi_{def} = \frac{4\pi}{\lambda} \Delta r = \frac{4\pi}{\lambda} \cdot T \cdot v + \phi_{non-linear} \tag{3}$$

where λ is the wavelength; r the range distance; b the normal baseline; θ the incidence angle; ε and v are the height error and linear velocity; T is the temporal baseline between both SAR acquisitions.

Before linear deformation retrieval, high coherence point targets are selected according to pixel's coherence stability by setting a suitable coherence threshold for the mean coherence image. Based on these point targets, differential phase are connected with Delaunay triangulation. Thus the phase slope between two neighbouring points $(x_m, y_m), (x_n, y_n)$ on an edge can be expressed as

$$\begin{aligned}\delta\phi_{dif}(x_m, y_m, x_n, y_n, T_i) = {} & \frac{4\pi}{\lambda} \cdot T_i \cdot [v(x_m, y_m) - v(x_n, y_n)] \\ & + \frac{4\pi}{\lambda} \cdot \frac{b(T_i)}{r(T_i)\sin(\theta_i)} \cdot [\varepsilon(x_m, y_m) - \varepsilon(x_n, y_n)] \\ & + [\beta(x_m, y_m) - \beta(x_n, y_n)] \\ & + [\alpha(x_m, y_m) - \alpha(x_n, y_n)] \\ & + [n(x_m, y_m) - n(x_n, y_n)]\end{aligned} \tag{4}$$

where (x_m, y_m) and (x_n, y_n) are pixel position coordinates; T_i is the time baseline of the ith interferogram; β the nonlinear component of velocity; α the atmospheric phase artefacts; and n the decorrelation noise.

It is assumed that, within the atmospheric correlation range 1-3km, the atmospheric phases are equal, thus the atmospheric components can be neglected. For the linear deformation velocity and height error are constants, thus the above phase slope can be modelled as

$$\begin{aligned}\delta\phi_{model}(x_m, y_m, x_n, y_n, T_i) = {} & \frac{4\pi}{\lambda} \cdot T_i \cdot [\Delta v_{model}(m,n)] \\ & + \frac{4\pi}{\lambda} \cdot \frac{b(T_i)}{r(T_i)\sin(\theta_i)} \cdot [\Delta\varepsilon_{model}(m,n)]\end{aligned} \tag{5}$$

where Δv are $\Delta\varepsilon$ velocity and height error increments, respectively. They can be retrieved by maximizing the following Ensemble Phase Coherence (EPC) (Ferretti, 2000)

$$\gamma_{model}(x_m, y_m, x_n, y_n) = \frac{1}{M} \cdot \left| \sum_{i=0}^{M} \exp \begin{bmatrix} j \cdot (\delta\phi_{dif}(x_m, y_m, x_n, y_n, T_i) \\ -\delta\phi_{model}(x_m, y_m, x_n, y_n, T_i)) \end{bmatrix} \right| \tag{6}$$

where j is the imaginary unit, M is the number of interferograms. When the maximum EPC is close to 1, the velocity and height error increments are close to the real value. Then, the linear velocity and height error on each point target is obtained by integrating Δv_{model} and $\Delta\varepsilon_{model}$ with EPC over 0.7 from a starting reference point.

To retrieve non-linear deformation, it is necessary to calculate the model phase contributed by linear deformation and height errors. By subtracting the model phase from differential phase, we get residual phases, which mainly include atmospheric phase, non-linear deformation component and phase noises. Phase noises can be reduced by spatial low pass filtering. Atmospheric phase and non-linear deformation can be separated according to their different frequency characteristics in temporal and spatial domains (Ferretti, 2000; Mora, 2003).

4. RESULTS

4.1 Results

50 Interferograms with baselines of less than 400m are generated. 3335 high coherent point targets are selected from the average coherence. Using small baseline DInSAR technique described above, average deformation velocity and accumulated deformation over point targets are generated. Deformation around the South Ring Expressway and the Wushu International Airport from 2006 to 2009 are analyzed. Figure 3 shows the average subsidence velocity on sparse points and Figure 4 shows the whole subsidence field by spatial interpolation.

It can be found that, along the South Ring Expressway, 3 sections with different subsidence rates can be separated from each other. The west section from A to B deforms with a velocity of 10 ~ -50 mm/a The middle one from B to C has the largest velocity ranging from -50 to -90 mm/a while the east section from C to D has the smallest velocity of around -15 mm/a. The middle section of the expressway is close to the Taiyuan Economic and Technological Development Zone in Xiaodian District where a large amount of factories are located. Over-exploration of ground water has led to the largest subsidence in this area. The west and east sections contain mainly rural areas where ground water demands are less, so the subsidence rates are much less.

As far as the Wushu International Airport is concerned, deformation rates are also spatially varied. The average deformation velocity in the southwest reaches to about -42 mm/a, while in the northeast it is -20 mm/a.

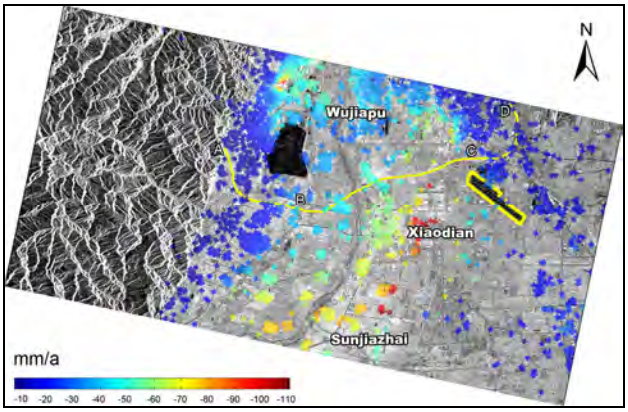


Figure 3. Average subsidence velocity in southern Taiyuan (2006-2009)

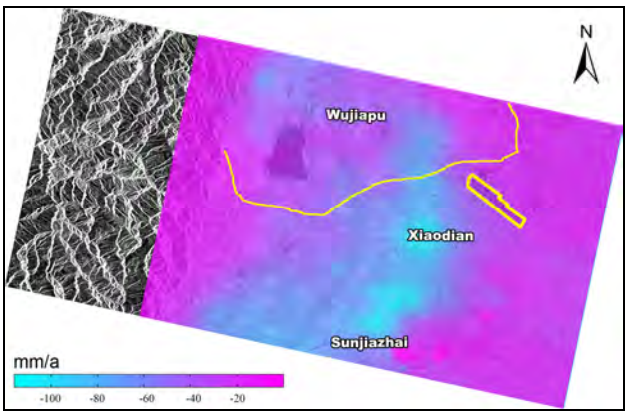


Figure 4. Subsidence field around expressway and airport (2006-2009)

4.2 Validation

To validate the DInSAR results, 3 levelling measurements, P1, P2 and P3 in the study area (Figure 1(b)), are used. Table 5 shows the comparison of the average deformation velocity between DInSAR results and levelling measurement. For the initial P3 measurement began in the end of 2006, the average deformation velocity at this point is not validated. From this table, it can be seen that the accuracy of DInSAR results at P1 and P2 is around -1.2 ~1.4 mm/a, demonstrating small baseline DInSAR technique can monitor ground deformation effectively. Besides the validation for average deformation velocity, we also compare the accumulative deformation on these 3 levelling points (Figure 6). The agreement between DInSAR results and levelling measurement is good.

Leveling points	DInSAR measurement	Leveling measurement	Errors
P1	-62.48	-61.33	-1.15
P2	-67.29	-68.67	1.38
P3	-63.25	-	-

Table 5. Comparison of average deformation velocity between DInSAR and levelling measurement (mm/a)

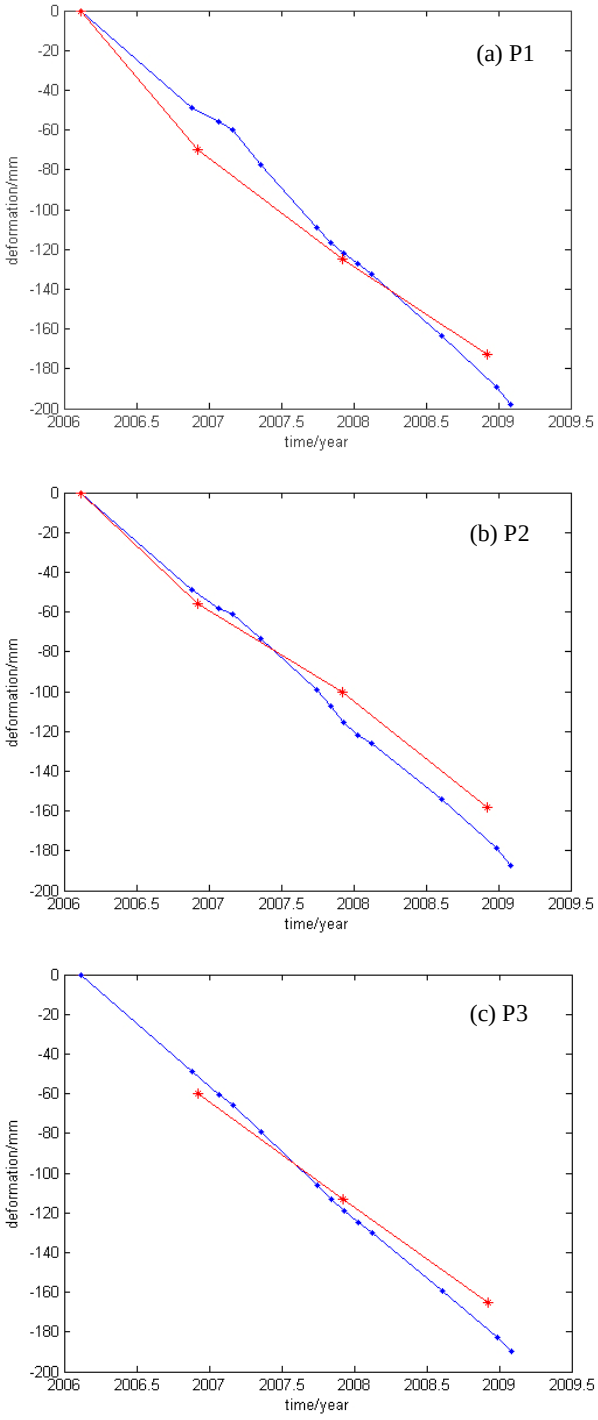


Figure 6. Comparison of accumulated deformation between DInSAR results (blue) and levelling measurements (red)

5. CONCLUSION AND DISCUSSION

Small baseline DInSAR technique combining the advantages of PS and CP methods has a great potential in deformation monitoring. Several characteristics of the technique can be summarized as the following: 1) it doesn't need as many SAR images as PS method; 2) it allows the generation of many small-

baseline multi-reference images to improve coherence; 3) it can derive linear and non-linear deformation fields at high coherence point targets.

We applied the technique to investigate ground deformation around the South Ring Expressway and Wushu International Airport in southern Taiyuan City. Along the South Ring Expressway, the maximum subsidence lies in the central section. The east section has the minimum subsidence and the west section has a medium subsidence velocity. Over the Wushu International Airport, the subsidence velocity is around -42 mm/a in the southwest and -20 mm/a in the northeast. The results suggest that the exploitation of Taiyuan Economic and Technological Development Zone in Xiaodian District has significant influence on local ground subsidence. Thus it is urgent to reduce ground water exploration in this area to control the subsidence.

REFERENCES

Berardino, P., Fornaro, G., Lanari, R., et al, 2002. A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms. *IEEE Trans Geosci Remote Sens*, 40(11), pp. 2375-2383.

Berardino, P., Casu, F., Fornaro, G., et al, 2003a. Small baseline DIFSAR techniques for earth surface deformation analysis. *FRING 2003 workshop*, Franscati, Italy.

Berardino, P., Fornaro, G., Lanari, R., et al, 2003b. A two-scale differential SAR interferometry approach for investigating earth surface deformations. *IGARSS 2003*, pp. 1184-1186.

Berardino, P., Casu, F., Lanari, R., et al, 2004. On the exploitation of the SBAS algorithm for the analysis of the deformations detected from the ERS and ENVISAT DInSAR data. *ERS-ENVISAT DInSAR Symposium*, 6–10 September 2004, Salzburg, Austria.

Casu. F., Manzo, M., Pepe, A., et al, 2008. SBAS-DInSAR Analysis of Very Extended Areas: First Results on a 60 000-km2 Test Site. *IEEE Trans. Geosci. Remote Sensing Letters*, 5 (3), pp. 438-442.

Ferretti, A., Rocca, F., Prati, C., 1999. Permanent scatterers in SAR interferometry. *IGARSS 1999*, Vol.3, pp.1528-1530

Ferretti, A., Prati, C., Rocca, F., 2000. Nonlinear subsidence rate estimation using permeanent scatterers in differential SAR interferometry. *IEEE Trans Geosci Remote Sens*, 28(5), pp. 2202-2212.

Ferretti, A., Prati, C., Rocca, F., 2001. Permanent scatterers in SAR interferometry. *IEEE Trans Geosci Remote Sens*, 39(1), pp. 8-20.

Mora, O., Mallorqui, J., Broquetas, A., 2003. Linear and nonlinear terrain deformation maps from a reduced set of interferometric SAR images, *IEEE Trans Geosci Remote Sens*, 41(10), pp. 2243-2253.

Sun, Z., Ma, T., Ma, J., et al, 2007. Effect of strata heterogeneity on spatial pattern of land subsidence in Taiyuan City. *Rock and Soil Mechanics*, 28 (2), pp. 399-408.

Vilardo, G., Ventura. G., Terranova. C., et al, 2009. Ground deformation due to tectonic, hydrothermal, gravity, hydrogeological, and anthropic processes in the Campania Region (Southern Italy) from Permanent Scatterers Synthetic Aperture Radar Interferometry. *Remote Sensing of Environment*, 113, pp. 197-212.

ACKNOWLEDGEMENTS

This research is supported by the National Key Basic Research and Development Program, China, under project number 2006CB701303, and Taiyuan Water Resources Bureau under the project, “Ground Subsidence Monitoring in Taiyuan Using Spaceborne InSAR Technique”. ENVISAT ASAR data are copyrighted and provided by European Space Agency (ESA). The authors would like to thank Dr. Zhong Lu (USGS) for his suggestive discussion and careful review. They also thank the Taiyuan Water Resources Bureau for the levelling data which validate DInSAR results in this research.

STRUCTURAL HIGH-RESOLUTION SATELLITE IMAGE INDEXING

Gui-Song Xia[1], Wen Yang[2], Julie Delon[1], Yann Gousseau[1], Hong Sun[2], Henri Maître[1]

[1]CNRS-LTCI, TELECOM ParisTech, 46 rue Barrault, 75013 Paris, France
e-mail: {*xia, delon, gousseau, henri.maitre*}*@enst.fr*
[2]Signal Processing Lab., School of Electronic Information, Wuhan University,
LuoJia Hill, Wuhan 430072, China
e-mail: {*yw, sh*}*@eis.whu.edu.cn*

KEY WORDS: retrieval and classification, high-resolution satellite image, structure, texture

ABSTRACT:

Satellite images with high spatial resolution raise many challenging issues in image understanding and pattern recognition. First, they allow measurement of small objects maybe up to 0.5 m, and both texture and geometrical structures emerge simultaneously. Second, objects in the same type of scenes might appear at different scales and orientations. Consequently, image indexing methods should combine the structure and texture information of images and comply with some invariant properties. This paper contributes to the indexing of high-resolution satellite images. We suggest a satellite image indexing method relying on topographic maps and a shape-based image indexing scheme. The proposed approach contains both the textural and structural information of satellite images and is also robust to changes in scale, orientation and contrast. Experimental analysis on a real satellite image database confirms the efficiency of the approach.

1 INTRODUCTION

Remote sensed satellite imaging has been widely applied to agriculture, geology, forestry, regional planning, and many other applications for analyzing and managing natural resources and human activities. In the past few years, with the development of imaging techniques, satellites with very high spatial resolution imaging systems have been launched, e.g. IKONOS, QuickBird, World-View-1, GeoEye-1, which enable satellite imagery to provide more accurate earth observation and measure small objects on the surface up to 0.5 m.

However, satellite images of high spatial resolution present many challenging problems in image understanding, information mining, and pattern recognition. First, with the enhancement of spatial resolution, more details on the earth surface emerge in satellite imagery. Unlike the case of low-resolution satellite images, where texture and intensity cues have been proved to be efficient for recognition (Li and Castelli, 1997, Richards and Jia, 2005, Ruiz et al., 2004), structures become more important for analyzing high-resolution satellite images. It is of great interest to investigate new image indexes, which can describe both the structure and texture information for high-resolution satellite image recognition. Second, in satellite images of high spatial resolution, objects contained in the same type of scenes might appear at different scales and orientations. For instance, the buildings in urban areas or the bridges on the river always show at various sizes and orientations. Moreover, if satellite images were taken under different weather conditions, there might be lighting changes between images of the same type. For these reasons, image indexing methods should comply with some invariant properties, such as scale invariance, orientation invariance and contrast invariance.

In order to extract structural features from optical satellite images of high-resolution, (Ünsalan and Boyer, 2004) proposed to use statistics of straight lines and their spatial arrangement over relatively small neighborhoods. (Bhattacharya et al., 2007, Bhattacharya et al., 2008) suggested to use geometrical information, e.g. *edge* and *Junction density*, from the extracted road network and segmented urban regions for structural satellite image indexing. As inspired by the works in computer vision, (Newsam and Yang, 2007) investigated interest point descriptors, such as *Scale-invariant feature transform* SIFT, for characterizing remote sensed images. Other structural features are computed from the pre-segmentation of images. One main disadvantage of this kind of approaches is that they rely on some pre-analysis of images, such as edge detection and segmentation, which are in themselves challenging problems. In addition, when these indexing schemes focus on structure information, they ignore the use of texture cues.

This paper contributes to structural indexing of high-resolution optical satellite images. The proposed indexing scheme is based on a complete morphological image representation, called *Topographical Map* (Caselles et al., 1999), which is made of all the connected components of the level lines of images. More precisely, the indexing of satellite images follows the shape-based indexing scheme, proposed by (Xia et al., 2009). First, satellite images are decomposed into tree of shapes, by using a morphological transformation, named *Fast Level Set Transformation* (FLST) (Monasse and Guichard, 2000). Then, image features are developed from those shape ensembles and their relationships. The derived shape-based features describe the structure distributions of images. They also encode the texture information of images, if taking shapes as *textons* (Zhu et al., 2005). Furthermore, the developed satellite images features are invariant to geometric transformations involving scaling and rotating and are robust to illumination changes.

The remainder of the paper is structured as follows. In Section 2, we give a brief review on our shape-based image indexing scheme. We then detail the proposed approach for high-resolution satellite images in Section 3. Section 4 provides the experimental analysis and Section 5 concludes the paper.

2 SHAPE-BASED IMAGE INDEXING FRAMEWORK

This section sketches the basics of our work, *i.e.* the topographic map and the shape-based image indexing framework.

2.1 Topographic map

For a gray-scale image u, the upper and lower level sets are defined respectively as $\chi_\lambda(u) = \{x \in \Omega;\ u(x) \geq \lambda\}$ and $\chi^\lambda(u) = \{x \in \Omega;\ u(x) \leq \lambda\}$, for $\lambda \in R$. The topographic map (Caselles et al., 1999) of the image u is made of the connected components of the topological boundaries of the upper level sets of the image (they could be equivalently defined from the lower level sets). Observe that the connected components of upper level sets (respectively of the lower level sets) are naturally embedded in a tree structure. Monasse and Guichard combined these two redundant tree structures, by drawing on the notion of *shape*, and developed an efficient way to compute a hierarchical representation of images (Monasse and Guichard, 2000), named FLST[1] as mentioned before. A ***shape*** is defined as the interior of a level line (the boundary of a level set). Figure 1 shows an example of the topographic map representation of a synthetic image.

It is shown that topographic map, the tree of shapes of an image has many impressive properties. First, it inherits a hierarchical structure from the nesting properties of level sets and it's a scale space without any geometrical degradation. Secondly, it's a complete image representation and can encode both the geometric and radiometric information simultaneously. And it's also invariant to any contrast changes.

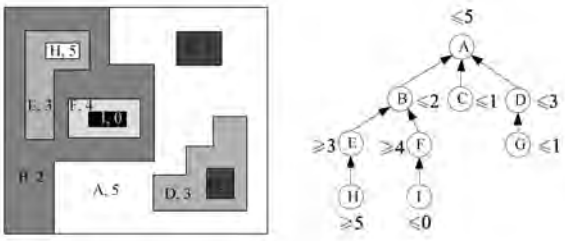


Figure 1: Representation of an image by its topographic map. Left: an original digital image; Right: representation of the image by its tree of shapes, where $(A, B, \ldots, I)$ denote the corresponding shapes.

2.2 Shape-based image indexing framework

By relying on the topographic map representation, Xia *et al.* proposed a shape-based invariant image indexing scheme in (Xia et al., 2009). A flowchart of the scheme is provided in Figure 2. The idea is to decompose images into shapes (by using FLST) and then develop image features from the shape ensembles and their relationships. It has been shown that the framework is very efficient for achieving geometric invariant texture features and obtain state-of-the-art performance on invariant texture recognition task.

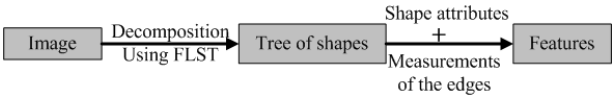


Figure 2: Shape-based image indexing framework.

As the topographic map provides a complete representation of images, the modeling of texture u is converted to the modeling of the tree of shapes (S, T), as $p(u) = p(T, S)$. The invariant texture features first rely on classical shape moments, then make use of the hierarchical structure of the topographic map.

[1]The codes of FLST are included in the free software MegaWave, and can be downloaded at http://megawave.cmla.ens-cachan.fr/.

The $(p+q)$th order central moments of shape s is defined as

$$\mu_{pq} = \int\int_s (x - \overline{x})^p (y - \overline{y})^q \, dxdy, \tag{1}$$

where $(\overline{x}, \overline{y})$ are the center of mass of s. According to the framework of (Xia et al., 2009), three different invariant features are developed from the invariant moments of shape ensembles S:

- EH: histogram of elongation $\epsilon = \frac{\lambda_2}{\lambda_1}$;
- CH: histogram of compactness $\kappa = \frac{1}{4\pi\sqrt{\lambda_1 \lambda_2}}$;
- CtH: histogram of contrast: $\gamma(x) = \frac{u(x) - \mathrm{mean}_{s(x)}(u)}{\sqrt{\mathrm{var}_{s(x)}(u)}}$.

where $s(x)$ is the smallest shape containing pixel x, λ_1 and λ_2 are the two eigenvalues of the normalized inertia matrix of shape s, with $\lambda_1 \geq \lambda_2$.

To develop features from the tree structure T, the parent-children relationships are used by defining an ancestor family of s, $\mathcal{N}_s^M$, as shapes containing the m-th ($m \leq M$) cascaded parents of s. The feature is:

- SH: histogram of scale ratio $\alpha(s) = \frac{\mu_{00}(s)}{\langle \mu_{00}(s') \rangle_{s' \in \mathcal{N}^M}}$;

where $\langle \cdot \rangle_{s' \in \mathcal{N}^M}$ is the mean operator on $\mathcal{N}^M$.

Observe that EH is invariant to similarity (translation, scaling and rotation) changes and CH, SH and CtH are invariant to affine transformations. Furthermore, all the four features are invariant to increasing contrast changes.

3 STRUCTURAL SATELLITE IMAGE INDEXING

This part devotes to the structural indexing of high-resolution satellite images, under the shape-based image indexing framework.

3.1 Structures of satellite imagery

Panchromatic images and multispectral images are the main types of optical satellite images acquired by optical remote sensing sensors. For satellite imagery in panchromatic format, all the structure information is, of course, contained in the gray scale image. However, for a multispectral satellite image $U = \{u_1, u_2, \ldots, u_L\}$ of L bands, an L-dimensional vector is stored for each pixel. In this case, we suppose that the main structure information of U is included by its p-energy channel $\mathcal{L}$, defined as

$$\mathcal{L} = \left(\sum_{u_i \in U} (u_i)^p \right)^{\frac{1}{p}}, \ p \geq 1. \tag{2}$$

Actually, Caselles *et al.* have proved that the main geometric information of natural color images are contained in their luminance channel (Caselles et al., 2002) (where $p = 1$). In the context of this work, as we shall see, we will only deal with natural color satellite images, so the analysis of structure information is based on the luminance channel. The same scheme could be applied to multispectral images using the p-energy channel $\mathcal{L}$.

3.2 High-resolution satellite image indexing

High-resolution satellite images allow to accurately represent small objects on the earth surface, such as cars, airplanes and buildings. An important and discriminative measurement for the objects is made of the shapes of their contour. For instance, bridges usually have elongated shapes, and the outlines of cars are usually compact and not too elongated.

First, following the shape-based image indexing scheme presented in Section 2, we also use the *elongation histogram* (EH), *compactness histogram* (CH), *scale ratio histogram* (SH), and *contrast histogram* (CtH) mentioned in Section 2 as indexes of satellite images.

Secondly, for the purpose of this paper, we propose to add several features, all invariant to similarity, a necessary invariance for satellite images. For instance, the orientation distribution of an image is also an available structure feature.
Orientation distribution: For a shape s, its orientation θ is calculated as

$$\theta = \frac{1}{2}\arctan\frac{2\mu_{11}}{\mu_{20}-\mu_{02}} \tag{3}$$

The resulted structure feature is the histogram of θ, named *Orientation Histogram* (OH), on all the shapes contained by the image.

We can also develop more structural features from the tree structure T, by using the M-order ancestor family $\mathcal{N}^M$:
Nested contrast: For a shape s, we define its nested contrast $\delta(s)$ as,

$$\delta(s) = \langle |u(s) - u(s')| \rangle_{s' \in \mathcal{N}^M}. \tag{4}$$

Maximum axis ratio: It's defined as ratio between the maximum eigenvalue of shape s and the average maximum eigenvalue among its ancestor family, as

$$\rho(s) = \frac{\lambda_1(s)}{\langle \lambda_1(s') \rangle_{s' \in \mathcal{N}^M}}. \tag{5}$$

where $\langle\cdot\rangle_{\mathcal{N}^M}$ is the mean operator on $\mathcal{N}^M$. We can similarly define the axis ratio by replacing λ_1 with λ_2. However, observe that $\mu_{00}(s) \propto \lambda_1\lambda_2$, so the similar definition involving to λ_2 will be redundant with the scale ratio $\alpha(s)$.
The corresponding features are the histograms of δ and ρ of all the shapes on the tree, called *nested contrast histogram* (NCH) and *maximum axis ratio histogram* (MAH), respectively. They are invariant to similar transformations.

input : Satellite image u
output: A set of indexes of u

Compute the luminance channel $\mathcal{L}(u)$ of the image;
Decompose $\mathcal{L}(u)$ into a tree of shapes $\{S, T\}$;
for *each* s_i *in* S **do**
    Compute the shape attributes $\epsilon_i, \kappa_i, \gamma_i, \theta_i$;
end
for *each* s_i *and* $\mathcal{N}^M_{s_i}$ *on* T **do**
    Compute α_i, δ_i, *and* ρ_i;
end
for *each attribute* ξ^i_k *in* $\{\epsilon_i, \kappa_i, \gamma_i, \theta_i, \alpha_i\, \delta_i, \rho_i\}$ **do**
    compute histogram $H(\xi_k) = \frac{\#\{\xi^i_k = \xi_k, 1 \le i \le N\}}{N}$.
end

Algorithm 1: Structural satellite image indexing.

A given high-resolution satellite image u is then characterize by 7 1-D histograms as detailed above (EH, CH, SH, CtH, OH, NCH, MAH). The pipeline of the indexing steps is given by Algorithm 1.

Remark that in order to compute the dissimilarity between two satellite images, we use the Kullback-Leibler divergence to compute distances between single descriptors and then add them together. Especially, for comparing two OH's, one is circularly shifted to compute the Kullback-Leibler divergence with the other and the minimum divergence among them is taken as the dissimilarity.

4 EXPERIMENTAL ANALYSIS AND DISCUSSION

In this part, we illustrate the proposed analysis scheme on high-resolution satellite image recognition tasks. In order to evaluate the efficiency of the proposed approach, we also compare it with other features for satellite image indexing.

As the structural image indexing approach proposed by (Bhattacharya et al., 2007) relies on edge or junction density in small patches of images, it is not invariant to scale changes and of course not comparable to our approach. The interest point based features, such as SIFT descriptor, might have some potential for the structural analysis of satellite images. For example, it has been used by (Newsam and Yang, 2007) for satellite image retrieval. However, one main disadvantage of interest points based approaches is that the detection of interest points depends on the contrast of the images. For some images where no contrasted structure presents, for instance, the meadow and forest displayed in Figure 3, there will be few interest point detected, which makes the indexing of those images difficult.

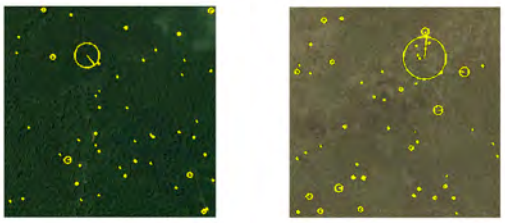


Figure 3: Two examples where few SIFT descriptors are detected. ***Left***: a forest image; ***Right***: a meadow sample.

In this work, we simply compare the proposed structural indexes with texture features based on Gabor filters with 6 scales and 8 orientations. In order to achieve rotation invariant, we average on all the 8 orientations.

4.1 High-resolution Database

To test the proposed satellite image indexing method, we collect a set of satellite images exported from Google Earth[2], which provides high-resolution satellite images up to 0.5 m. Some samples of the database are displayed in Figure 4.1. (The database can be downloaded at (Xia, 2009).) It contains 12 classes of meaningful scenes in high-resolution satellite imagery, including *Airport, Bridge, River, Forest, Meadow, Pond, Parking, Port, Viaduct, Residential area, Industrial area, and Commercial area*; For each class, there are 50 samples. It's worth noticing that the image samples of the same class are collected from different regions in satellite images of different resolutions and then might have different scales, orientations and illuminations.

[2] http://earth.google.com/

Figure 4: Some samples of the testing high-resolution satellite image database. For each class, there are 50 samples, and 4 of which are shown here.

4.2 Retrieval and classification

We apply the proposed analysis scheme to two common satellite image recognition tasks: retrieval and classification. For retrieval, one sample is used as a query image (thus removed from the database) and the N_r most similar samples are retrieved from the database. One after the other, all samples in the database are used as query images, and the average recall is computed in function of the number N_r of retrievals for evaluating the performance.

In the classification experiment, N_c samples from each class are randomly chosen as a training set and the remaining samples are classified thanks to a nearest-neighbor classifier. The rate of correct classifications is then computed as a function of N_c. In order to consolidate the results, classification rates are averaged on a sequence of 200 random training sets.

Figure 5(a) shows the average retrieval performance on the whole database. It indicates that by using only one sample, averagely 52.84% samples of the same class can be correctly retrieved among the first 49 matches. However, in the same context, if we use the mean and standard deviation of Gabor filter responses, only 21.19% samples can be retrieved averagely. According to the performance curve, when the number of matches is extended to 200, 90.50% samples can be retrieved by using the structural indexes. But the same percentage to Gabor features is 54.13%.

Some illustrations of the retrieval results are displayed in Figure 6, 8(a), 8(b), where a query image is followed by its first 49 closest samples. The retrieval results for all samples can be found at (Xia, 2009).

Figure 6 shows a retrieval result of bridge category, which is very structured. Observe that even though there are large illumination changes between samples and the query image, the method works well, thanks to the contrast invariance of the indexing scheme. It's also interesting to inspect the false alarms and observe that there often contains some structures similar to bridges, see the parts framed in red inside Figure 7. Figure 8(a) and 8(b) illustrate two retrieval examples respectively on river and viaduct class. Even structures in this two classes are complicated, the proposed approach works well.

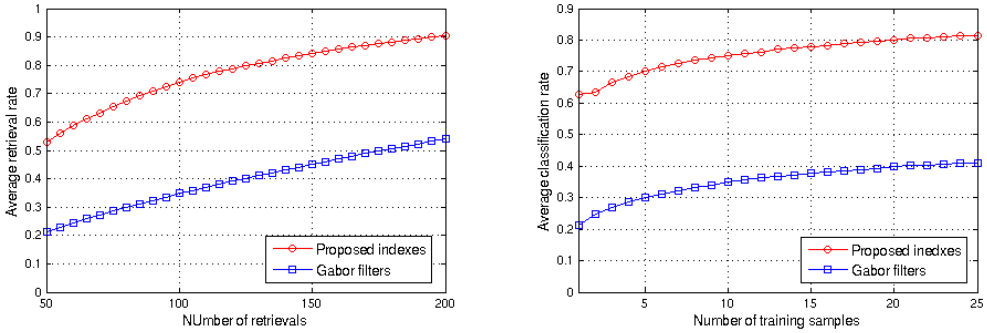
Figure 5(b) shows the average classification performance by using nearest-neighbor classifier, when the number of training images ranges from 1 to 25. It indicates that the structural indexes outperforms the Gabor features dramatically. Furthermore, Table 1 shows the average classification rate for each class of the database. We can see that Gabor features are efficient only on some texture classes, e.g. forest and meadow. The proposed structural indexes work well on classes with complicated structures such as viaduct and airport, and also on classes containing more textures.

However, we found that the structural indexes can not distinguish industrial and residential classes well. This is because those two categories share many similar structures, and some semantic information of the scene might be helpful (Bordes and Maître, 2007).

Figure 7: Some false alarms of bridge retrievals. The parts framed in red really contain some bridge-like structures.

Category	1 training sample		25 training sample	
	StructInd	GaborF	StructInd	GaborF
Airport	53.39	12.72	82.31	39.93
Bridge	40.45	7.73	83.42	27.57
Commercial	48.74	23.17	82.02	43.06
Forest	80.94	42.54	94.62	74.45
Industrial	42.93	20.11	78.37	37.75
Meadow	75.06	35.08	95.45	55.01
Parking	68.74	10.58	81.92	24.89
Pond	63.96	29.91	82.95	48.09
Port	50.93	12.79	73.80	26.66
Residence	32.17	19.04	48.87	40.07
River	60.23	27.28	88.66	49.79
Viaduct	64.19	13.24	86.76	23.78

Table 1: Average classification rate (%) of Structural indexes (using **StructInd** for short) and Gabor Features (using GaborF for short) on each category of the database, with the number of training samples as 1 and 25, respectively.



(a) retrieval performance (b) classification performance

Figure 5: Average retrieval (a) and classification (b) performance.

Figure 6: A retrieval result of the bridge category obtained by using the proposed indexing scheme. The query image is in the first position and the 49 most similar samples follow, ordered by their matching scores. The false samples are framed in red.

5 CONCLUSION

In this paper, we have developed some structural features for indexing high-resolution satellite image, based on the topographic map and under the shape-based image indexing framework. The experimental analysis shows that the indexes can balance the structures and textures information in high-resolution satellite images and provide impressive image recognition performances.

However, we should observe that we simply adopted the nearest-neighbor classifier for classification. The recognition performance, of course, benefits from some more powerful classification scheme, e.g. SVM.

ACKNOWLEDGEMENTS

This work was partly supported by a grant from the National Natural Science Foundation of China (No.40801183, 60890074).

REFERENCES

Bhattacharya, A., Roux, M., Maître, H., Jermyn, I. H., Descombes, X. and Zerubia, J., 2007. Computing statistics from man-made structures on the earth's surface for indexing satellite images. International Journal of Simulation Modelling 6(2), pp. 73–83.

Bhattacharya, A., Roux, M., Maître, H., Jermyn, I. H., Descombes, X. and Zerubia, J., 2008. Indexing of mid-resolution satellite images with structural attributes. In: Proc. 21th ISPRS Congress, Commission IV, Beijing, China, pp. 187–192.

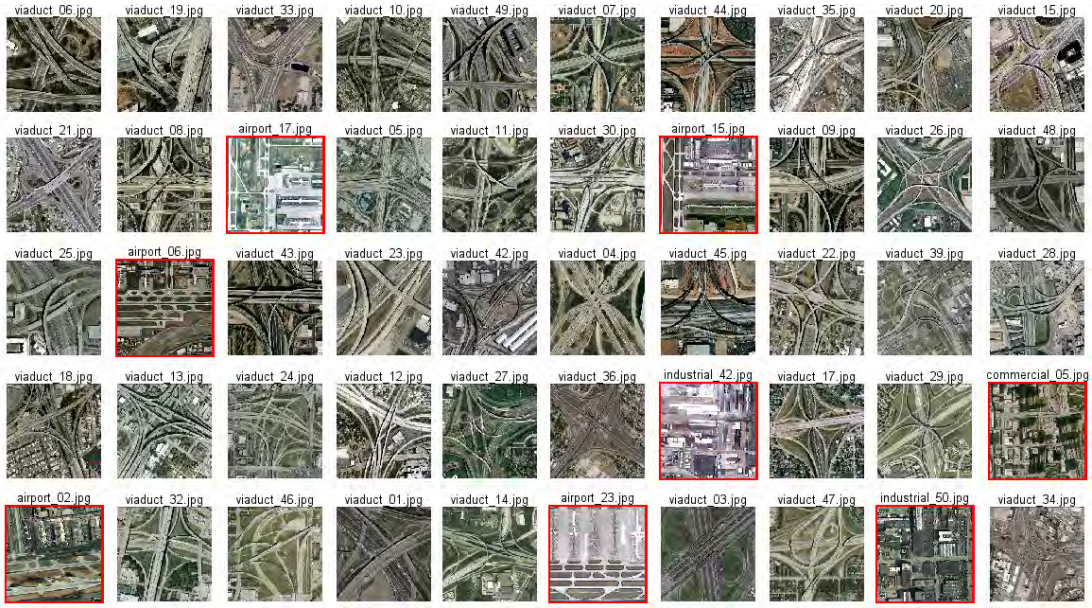
Bordes, J.-B. and Maître, H., 2007. Semantic annotation of satellite images. In: MLDM Posters, pp. 120–133.

Caselles, V., Coll, B. and Morel, J.-M., 1999. Topographic maps and local contrast changes in natural images. International Journal of Computer Vision 33(1), pp. 5–27.

Caselles, V., Coll, B. and Morel, J.-M., 2002. Geometry and color in natural images. J. Math. Imaging Vis. 16(2), pp. 89–105.

Li, C.-S. and Castelli, V., 1997. Deriving texture feature set for content-based retrieval of satellite image database. In: Proceed.

(a) A retrieval example of river category



(b) A retrieval example of viaduct category

Figure 8: Some retrieval results obtained on the database using the proposed indexing scheme. The query image is in the first position and the 49 most similar samples follow, ordered by their matching scores. The false samples are framed in red.

International Conference on Image Processing, IEEE Computer Society, Washington, DC, USA, p. 576.

Monasse, P. and Guichard, F., 2000. Fast computation of a contrast invariant image representation. IEEE Trans. Image Processing 9(5), pp. 860–872.

Newsam, S. and Yang, Y., 2007. Comparing global and interest point descriptors for similarity retrieval in remote sensed imagery. In: GIS '07: Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems, ACM, New York, NY, USA, pp. 1–8.

Richards, J. A. and Jia, X., 2005. Remote Sensing Digital Image Analysis: An Introduction. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Ruiz, L. A., Fdez-Sarria, A. and Recio, J. A., 2004. Texture feature extraction for classification of remote sensing data using wavelet decomposition: a comparative study. In: Proc. 20th ISPRS Congress, Commission IV, pp. 1109–1114.

Ünsalan, C. and Boyer, K. L., 2004. Classifying land development in high resolution panchromatic satellite images using straight line statistics. IEEE Trans. Geosci. Remote Sens 42, pp. 907–919.

Xia, G.-S., 2009. High-resolution satellite image indexing project webpage. http://www.tsi.enst.fr/~xia/satellite_image_project.html.

Xia, G.-S., Delon, J. and Gousseau, Y., 2009. Shape-based invariant texture indexing. International Journal of Computer Vision, to appear.

Zhu, S.-C., Guo, C.-E., Wang, Y. and Xu, Z., 2005. What are textons? International Journal of Computer Vision 62(1-2), pp. 121–143.

MULTISTAGE ALGORITHM FOR LOSSLESS COMPRESSION OF MULTISPECTRAL REMOTE SENSING IMAGES

A. Zamyatin

Tomsk Polytechnic University 30, Lenina ave., Tomsk, 634050, Russia - zamyatin@tpu.ru

KEY WORDS: Multispectral, Compression, Algorithms, Image, Comparison

ABSTRACT:

A preliminary comparison between loss and lossless compression approaches for the remote sensing data processing was made. A three-stage lossless compression algorithm of multispectral remote sensing images based on wavelet transformations and intra-bands correlation is proposed and developed. It allows one to consider peculiarities of remote sensing data and to increase the compression ratio of the algorithm. The paper describes a modification of the compression algorithm aimed at considerable improvement of computational performance and based on bands trimmed enumeration and data selective use. A research of the three-stage algorithm performance was carried out in comparison with the universal compression algorithms such as *WinRar*, *WinZip* and *JPEG2000* using data from various remote sensing systems showing to some extent a superiority in the compression ratio, as well as some insignificant lag of the computational performance was identified.

1. INTRODUCTION

Due to the constantly improving technical features of remote sensing (RS) systems and the extended use of RS data for solving various tasks, the data amount handled by RS modern systems is in terabytes and it continues to increase steadily. Hence, the resolution of RS data compression problems using different approaches, software and hardware aimed at increasing the effectiveness of data processing, storage and transmission along communication channels is becoming more relevant (Cagnazzo etc., 2006; Marcellin etc., 1995; Motta etc., 2006; Salmon, 2007; Ziv & Lempel, 1977). In general, data compression approaches can be presented by loss and lossless ways. Let us consider first a possibility to use the lossless approach for RS data compression.

Lossy algorithms are widely used for image compression tasks and are characterized by high possible compression ratio (Jacquin, 1993). For instance, the most popular algorithms are based on fractal approach and are used both for color and grayscale images without a sharp color change (for example, photo pictures). It allows one getting high compression rates over 200. The fractal algorithm might be implemented for each RS band of $N \times M$ size and it is based on the so called *domain* and *rang* regions (areas) processing (Fig. 1).

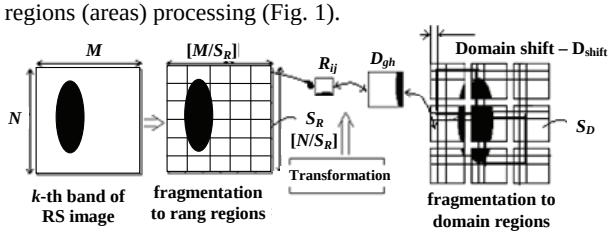


Figure 1. A general scheme of the fractal algorithm

The fractal algorithm could bring a significant compression ratio for the large images that are common for RS images. The rang regions are non-overlapping image fragments, the sum of which fully covers the image. Domain regions are similar to the rang regions. They are also formed by an image decomposition to some equivalent grid but differ in a bigger size and a shift. Detailed description of fractal algorithm for image compression can be found in (i.e. Jacquin, 1993), but briefly it could be generally presented in the following way:

1. Decompose the image by equivalent grid to the same set of square rang regions R_{ij} with S_R number of elements (pixels) where $i = 0, S_R, \ldots, ([M/S_R] - 1) \times S_R$; $j = 0, S_R, \ldots, ([N/S_R]-1) \times S_R$.
2. Form D_{gh} domain regions with S_D dimension, where $g = 0, D_{shift}, 2 \times D_{shift}, \ldots, [(M - S_D)/D_{shift}] \times D_{shift}$; $h = 0, D_{shift}, 2 \times D_{shift}, \ldots, [(N - S_D)/ D_{shift}] \times D_{shift}$.
3. For every R_{ij} element search through all D_{gh} elements and form $R_{ij} \rightarrow D_{gh}$ affine transformation:

$$f_k(x, y, I(x,y)) = \begin{pmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & p \end{pmatrix} \times \begin{pmatrix} x \\ y \\ I(x,y) \end{pmatrix} + \begin{pmatrix} u \\ v \\ q \end{pmatrix},$$

where $k = 1, 2, \ldots, k_{af}$ – the number of affine transformation, x,y – coordinates of image pixel, $I(x,y)$ – pixel brightness with (x,y) coordinates, p – a coefficient that is called brightness (operating with grayscale image $p = 1$), a, b, c, d – coefficients oriented on performing the rotation and reflecting symmetry, u,v – shift of the point with (x,y) coordinates along the axes, q – coefficient that is called the shift of brightness point with (x,y) coordinates and is calculated as:

$$q = [\sum_{m=1}^{S_R} \sum_{l=1}^{S_R} r_{ml} - \sum_{m=1}^{S_R} \sum_{l=1}^{S_R} d_{ml}] / n^2,$$

where m, l – pixel coordinates, $r_{ml} - I(m, l)$ in R_{ij} rang region, $d_{ml} - I(m, l)$ in D_{gh} domain region.

4. Choose from all transformations one with the least error, calculated as:

$$d = \sum_{m=1}^{S_R} \sum_{l=1}^{S_R} [(r_{ml} - d_{ml}) - q]^2.$$

5. Fix the number of k_{af}, coordinates of the top-left corner of R_{ij} rang region, q shift of brightness in file.

The described fractal algorithm consequently implements the compression for every band of a multispectral RS image.

To evaluate the application outlook of the proposed lossy compression algorithm it is implemented in a framework of the fractal approach, some test experiments were performed. They revealed that the loss compression algorithm could provide

rather high compression ratios. A visual evaluation comparison between the initial and the uncompressed RS images (example on Fig. 2a,b) demonstrated a possibility to use the uncompressed RS images for the manual visual interpretation in some cases.

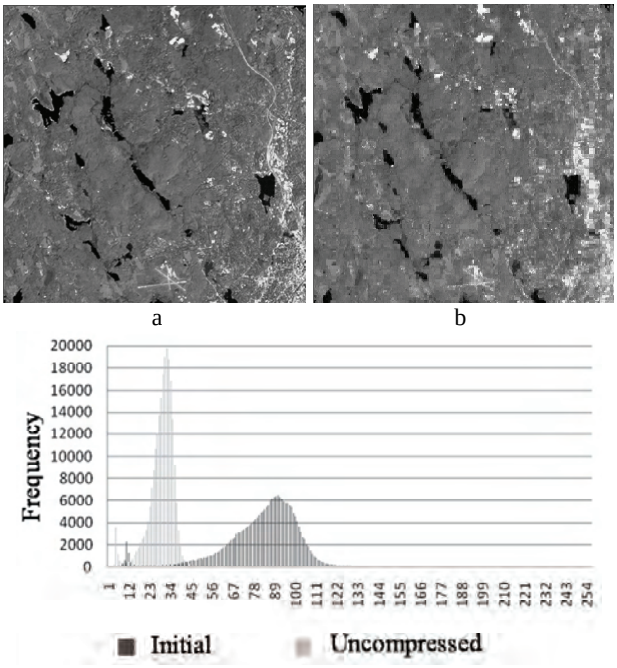


Figure 2. Results of lossy compression: original multispectral RS (a), uncompressed RS image (b), histograms of a RS band (c)

At the same time a comparison of the frequency brightness histograms proves the significant changes in statistic brightness characteristics of the uncompressed RS image in comparison with the initial RS one (Fig. 2c). That implies possible difficulties in further automatic processing and classification of these uncompressed RS data, though high compression rates could be provided.

Thus, taking into account a possible demand in RS data preliminary processing and automated classification it is more reasonable to avoid distortion of statistical albedo characteristics of uncompressed RS images. It is possible to avoid with lossless compression algorithms which are considered most valuable for the RS data processing.

There are two conceptually different approaches to the RS data lossless compression. One approach implies the use of the universal and well-known compression algorithms do not take into account the specific features of RS data and which is represented in such software as *WinRar*, *WinZip* or which uses the compression standard of black-and-white and colored images as *JPEG2000* (Christopoulos, 2001; Taubman & Marcellin, 2002; Salmon, 2007;). Another approach is focused on new compression algorithms which take into account not only bands data as simple black-and-white images, but also relation between RS image bands. In spite of the fact that such approach is more complex from computational point of view its application allows one to achieve considerable compression ratio due to specific features of RS data.

In this respect the paper is aimed at the development and research of multispectral algorithms of lossless RS images compression based on both independent data processing in different bands and considering their intra-bands correlation to improve the compression ratio in comparison with well-known universal compression algorithms.

2. THREE-STAGE COMPRESSION ALGORITHM

The wavelet transformations allow one to obtain coefficients which can be compressed significantly better than initial image data. This approach to the lossless image compression is considered as the most efficient one (Salmon, 2007; Christopoulos, 2001). Multispectral RS images represent albedo values obtained in different spectral bands and which as a rule have considerable intra-bands correlation. If such functional relation is known it is possible to considerably decrease data transformation range using deviation (difference) values between the functional relation and the actual initial values. This will allow one to use significantly less number of bits to store such deviations than it is necessary to store the initial RS data. Eventually this allows one to increase the final compression ratio.

The application of the abovementioned transformations in the compression allows one to use the advantages of the wavelet transformations and the existing relation between the bands of multispectral RS images. Taking this into account, the compression algorithm can be carried out in three stages:

- to carry out wavelet transformation of initial data and obtaining corresponding transformation coefficients;
- to consider functional relation of albedo values between different image bands and to form a set of data deviations;
- to compress obtained data using one of the traditional compression algorithms.

Let us consider the stages of the suggested three-stage compression algorithm of multispectral RS images in detail.

The wavelet transformation is applied to an initial RS image according to rows and columns with given number of levels thus setting high-frequency and low-frequency components (Fig. 3).

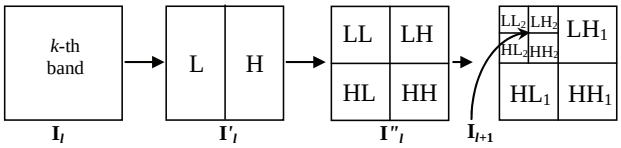


Figure 3. High-frequency and low-frequency components in wavelet transformation from RS image *k*-band data

To describe the first stage of compression algorithm step-by-step it is necessary to identify the following notational conventions: M – number of rows, N – number of columns, K – number of initial multispectral RS bands, l – an index of the current transforming level, L – number of transforming levels, sign "$\lfloor\ \rfloor$" – rounding up to the integer number.

Step 1. Set $m = 0$, $n = 0$, $k = 1$, $l = 1$.

Step 2. Obtain even $\mathbf{I}_l[m, 2j, k]$ and uneven $\mathbf{I}_l[m, 2j+1, k]$ components at $j = 0, 1, \ldots, \lfloor N/2^{l-1} \rfloor$ from the initial image $\mathbf{I}_l[m, n, k]$.

Step 3. Calculate low-frequency $\mathbf{Y}[m, 2j]$ and high-frequency $\mathbf{Y}[m, 2j+1]$ components by using 5 and 3 summand components of the initial image $\mathbf{I}_l$, at $j = 0, 1, \ldots, \lfloor N/2^{l-1} \rfloor$:

$\mathbf{Y}[m, 2j] = (-\mathbf{I}_l[m, 2j-1, k] + 2\cdot\mathbf{I}_l[m, 2j, k] + 6\cdot\mathbf{I}_l[m, 2j+1, k] + 2\cdot\mathbf{I}_l[m, 2j+2, k] - \mathbf{I}_l[m, 2j+3, k])/8$,

$\mathbf{Y}[m, 2j+1] = (-\mathbf{I}_l[m, 2j, k] + 2\cdot\mathbf{I}_l[m, 2j+1, k] - \mathbf{I}_l[m, 2j+2, k])/2$ or after approximation and rounding up operations

$\mathbf{Y}[m, 2j] = \mathbf{I}_l[m, 2j, k] + \lfloor(\mathbf{Y}[m, 2j-1] + \mathbf{Y}[m, 2j+1]) + 2)/4\rfloor$,

$\mathbf{Y}[m, 2j+1] = \mathbf{I}_l[m, 2j+1, k] - \lfloor(\mathbf{I}_l[m, 2j, k] + \mathbf{I}_l[m, 2j+2, k])/2\rfloor$

Step 4. If $m < M$ then $m = m + 1$, step 2, otherwise step 5.

Step 5. Form the image $\mathbf{I}'_l$, containing high-frequency and low-frequency areas (according to columns):

For $p = 0, 1,\ldots, \lfloor M/2^{l-1} \rfloor$, $q = 0, 1,\ldots, \lfloor N/2^{l} \rfloor$, $\mathbf{I}'_l[p, q, k] = \mathbf{Y}[m, 2j]$;
For $p = 0, 1,\ldots, \lfloor M/2^{l-1} \rfloor$, $q = \lfloor N/2^{l} \rfloor + 1, \lfloor N/2^{l} \rfloor + 2,\ldots, \lfloor N/2^{l-1} \rfloor$, $\mathbf{I}'_l[p, q, k] = \mathbf{Y}[m, 2j+1]$.
Step 6. Calculate low-frequency $\mathbf{Y}[m, 2j]$ and high-frequency $\mathbf{Y}[m, 2j + 1]$ components on the basis of $\mathbf{I}'_l$:
$\mathbf{Y}[2j, n] = \mathbf{I}'_l[2j, n, k] + \lfloor(\mathbf{Y}[2j\text{-}1, m] + \mathbf{Y}[2j + 1, n]) + 2)/4\rfloor$,
$\mathbf{Y}[2j + 1, n] = \mathbf{I}'_l[2j+1, n, k] - \lfloor(\mathbf{I}'_l[m, 2j, k] + \mathbf{I}'_l[m, 2j + 2, k])/2\rfloor$.
Step 7. If $n < N$ then $n = n + 1$, step 6, otherwise step 8.
Step 8. Form the image containing high-frequency and low-frequency areas (according to rows):
$\mathbf{I}''_l[p, q, k] = \mathbf{Y}[2j, q]$ at $p = 0, 1,\ldots, \lfloor M/2^{l-1} \rfloor$ for $q = 0, 1,\ldots, \lfloor N/2^{l-1} \rfloor$;
$\mathbf{I}''_l[p, q, k] = \mathbf{Y}[2j+1, q]$ at $p = \lfloor M/2^{l-1} \rfloor + 1, \lfloor M/2^{l-1} \rfloor + 2, \ldots, \lfloor M/2^{l} \rfloor$ for $q = 0, 1,\ldots, \lfloor N/2^{l-1} \rfloor$.
Step 9. For $p = 0, 1,\ldots, \lfloor M/2^{l-1} \rfloor$ and $q = 0, 1,\ldots, \lfloor N/2^{l-1} \rfloor$ form $\mathbf{I}_l^w[p, q, k] = \mathbf{I}''_l[p, q, k]$, if $l < L$ then $l = l + 1$, step 2, otherwise step 10.
Step 10. If $k < K$ then $k = k + 1$, step 2, otherwise step 11.
Step 11. End.

The result of this stage is the image $\mathbf{I}_l^w[m, n, k]$ containing obtained low-frequency and high-frequency components found on the basis of the initial image $\mathbf{I}[m, n, k]$ using wavelet transformation with L levels of depth.

The main point of the second stage is to consider band-to-band correlation by defining the deviations between an obtained functional relation (of the 1st order in the given case) and the actual values in corresponding bands data obtained at the first stage by the wavelet transformation. Storage and further processing of deviations (not the initial data) is characterized by considerably low value change range, which requires the less number of bits and will allow one to compress such data with the higher ratio. The diagram of such transformation is shown in Fig. 4.

Given that $\mathbf{I}_l^Q[m, n, k]$ – matrix of $\mathbf{I}_l^w[m, n, k]$ image quadrant values with index Q, $Q = \{\text{LL, LH, HL, HH}\}$, $\boldsymbol{e}_l[m, n, k]$ – deviation matrix, then step-by-step description of the compression algorithm second stage can be presented in the following way:
Step 1. For $m = 0,1,\ldots,M$, $n = 0,1,\ldots,N$, $Q = \{\text{LL},\ldots, \text{HH}\}$,
$\boldsymbol{e}_L^Q[m, n, 1] = \mathbf{I}_L^Q[m, n, 1]$,
$\boldsymbol{e}_L^{LL}[m, n, 2] = \mathbf{I}_L^{LL}[m, n, 2] - \mathbf{I}_L^{LL}[m, n, 1]$.
Step 2. For $m = 0, 1,\ldots, \lfloor M/2^{l} \rfloor$, $n = 0, 1,\ldots, \lfloor N/2^{l} \rfloor$, $k = 3,4,\ldots,K$, $\boldsymbol{e}_L^{LL}[m,n,k] = \mathbf{I}_L^{LL}[m,n,k] - 2\cdot\mathbf{I}_L^{LL}[m, n, k–1] + \mathbf{I}_L^{LL}[m, n, k –2]$.
Step 3. Calculate coefficients for L level:
$\mathbf{w}_L^{LH}[k] = (\mathbf{J}_L^{LL}[k]^T\times\mathbf{J}_L^{LL}[k])^{-1}\times\mathbf{J}_L^{LL}[k]^T\times\mathbf{J}'_L{}^{LL}[k]$, $\mathbf{J}_L$ – m by n matrix $Z\times(\lfloor M/2^{L-1}\rfloor+1)\times(\lfloor N/2^{L-1}\rfloor+1)$, where $Z = 2$ at $k<2$, otherwise $Z = 1$.
$\mathbf{J}_L^{LL}[k] = [[\mathbf{I}_L^{LL}[0, 0, k\text{-}1], [\mathbf{I}_L^{LL}[0, 1, k\text{-}1], \ldots, \mathbf{I}_L^{LL}[\lfloor M/2^{L-1}\rfloor, \lfloor N/2^{L-1}\rfloor, k\text{-}1], [\mathbf{I}_L^{LL}[0, 0, k\text{-}2], [\mathbf{I}_L^{LL}[0, 1, k\text{-}2], \ldots, \mathbf{I}_L^{LL}[\lfloor M / 2^{L-1}\rfloor, \lfloor N/2^{L-1}\rfloor, k\text{-}2]]$, $\mathbf{J}'_L{}^{LL}[k]$ – vector with length $(\lfloor M/2^{L}\rfloor+1)\times(\lfloor N/2^{L}\rfloor + 1)$, $\mathbf{J}'_L{}^{LL}[k] = [[\mathbf{I}_L^{LL}[0, 0, k], [\mathbf{I}_L^{LL}[0, 1, k], \ldots, \mathbf{I}_L^{LL}[\lfloor M/2^{L-1}\rfloor, \lfloor N/2^{L-1}\rfloor, k]]$, where "T" – matrix transforming character, "–1" – inverse matrix character.
Step 4. For $l = L,L\text{-}1,L\text{-}2,\ldots,1$ calculate other vectors of coefficients $\mathbf{w}_l^{LH}$ and coefficients $\mathbf{w}_l^{HL}$, $\mathbf{w}_l^{HH}$:
$\mathbf{w}_l^{LH}[k] = [(\mathbf{J}_{l-1}^{HH}[k]^T\times\mathbf{J}_{l-1}^{HH}[k])^{-1}\times\mathbf{J}_{l-1}^{HH}[k]^T\times\mathbf{J}'_{l-1}{}^{HH}[k])]$,
$\mathbf{w}_l^{HL}[k] = [(\mathbf{J}_l^{LH}[k]^T\times\mathbf{J}_l^{LH}[k])^{-1}\times\mathbf{J}_l^{LH}[k]^T\times\mathbf{J}'_l{}^{LH}[k])]$,
$\mathbf{w}_l^{HH}[k] = [(\mathbf{J}_l^{HL}[k]^T\times\mathbf{J}_l^{HL}[k])^{-1}\times\mathbf{J}_l^{HL}[k]^T\times\mathbf{J}'_l{}^{HL}[k])]$.
Step 5. For $l = 1,2,\ldots,L$ and $Q = \{\text{LH, HL, HH}\}$ find

$$\boldsymbol{e}_l^Q[m,n,k] = \mathbf{I}_l^Q[m,n,k] - (\mathbf{w}_l^Q[k])^T \times \begin{pmatrix} \mathbf{I}_l^Q[m,n,k-1] \\ \mathbf{I}_l^Q[m,n,k-2] \end{pmatrix}.$$

Step 6. If $k < K$ then $k = k + 1$, step 4, otherwise step 7.
Step 7. End.

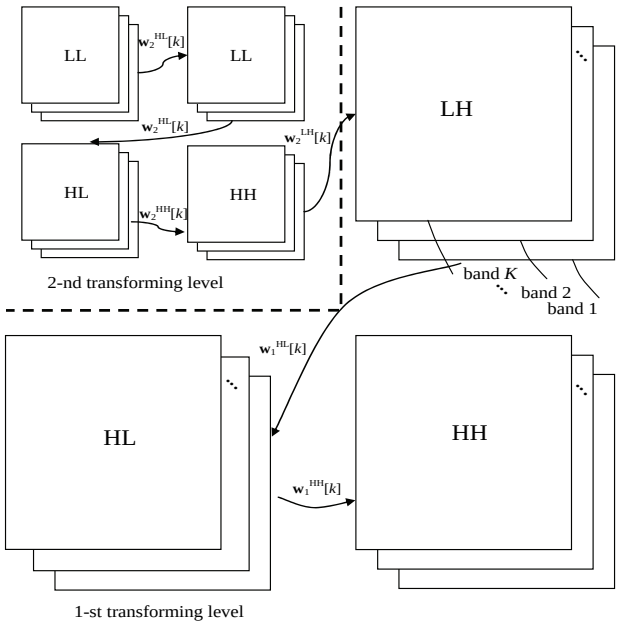


Figure 4. Generalized scheme of intra-bands correlation for transforming levels $l = 1$ and $l = 2$

The result of the second stage is deviation matrix $\boldsymbol{e}_l[m,n,k]$, which can be compressed by some algorithm at the final stage. In this case it is suggested to use a well-known arithmetical algorithm to compress the obtained data (Salmon, 2007; Witten etc., 1987).

To form the uncompressed multispectral image $\mathbf{I}[m,n,k]$ out of $\boldsymbol{e}_l[m,n,k]$ it is necessary to carry out a number of transformations opposite the above mentioned ones.

3. EXPERIMENTS

To define the efficiency of the suggested three-stage algorithm from the point of view of compression ratio and computational performance, as well as its validity limits, numerous experiments using multispectral RS images of different RS data systems (Table 5) in data format of raster geoinformation system *Idrisi Kilimanjaro* were carried out together with their comparison with experimental results obtained for well-known prototypes – *WinRar*, *WinZip* and *JPEG2000* of *FastStone Image Viewer* (Salamon, 2007; Kiely etc., 2006; Christopoulos etc., 2001). The algorithm was implemented in *Borland Developer Studio 2006* without special attention to the code optimization. That is why there are spaces for possible improvements. Experiments were carried out on *Intel Pentium IV* PC, 2.8 GHz, 1 Gb memory under *Windows XP* (SP 3).

No.	RS data systems	Number of bands	Size of image, pixels	Size of file, bytes
1	SPOT	3	509 × 571	871917
2	SPOT	3	615 × 558	1029510
3	ADAR–5000	3	541 × 440	714120
4	Airphoto	3	652 × 694	1357464
5	Landsat–MSS	4	558 × 560	1249920
6	Landsat–MSS	4	480 × 480	921600
7	Landsat–TM	6	934 × 700	3922800
8	Landsat–TM	7	500 × 500	1750000
9	Landsat–TM	7	525 × 280	1029000
10	Flightline C1	12	949 × 220	2505360

Table 5. RS experimental data

As was mentioned above, one of the key parameters of the wavelet transformation is its L depth which can be set over in a wide range. The increase of L value might lead to an increase of the compression ratio due to making greater high-frequency area. To define the most acceptable range of L parameter a number of experiments were carried out, the results of which are shown in Fig. 3.

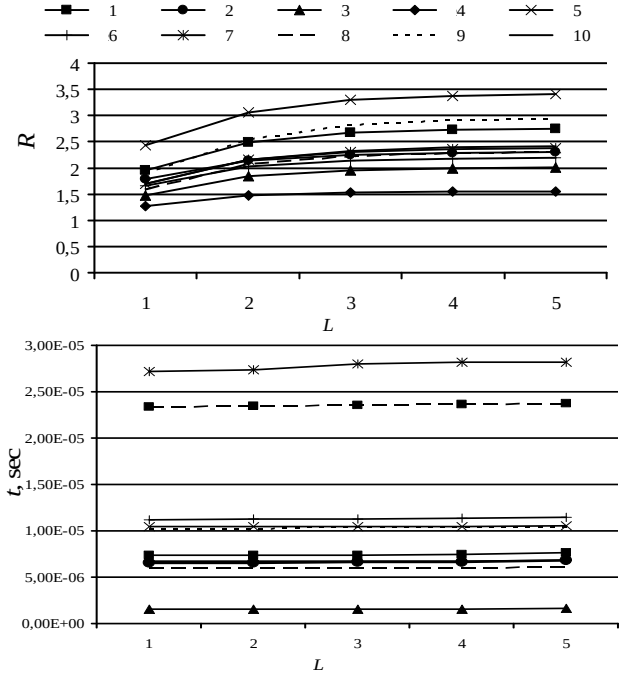


Figure 6. Relation of compression ratio and time on wavelet transformation with depth L

Experimental results presented in Fig. 6 show that although algorithm computational performance for $L\in[1;5]$ does not depend much on the transformation depth the compression ratio R ceases to increase even at $L\geq3$. Due to this it is advisable to accept here $L = 3$.

It is obvious that at the result of the second stage of the algorithm are the deviations $\boldsymbol{e}_l$ obtained with functional relation (set by coefficients $\mathbf{w}_l^{LH}$, $\mathbf{w}_l^{HL}$, $\mathbf{w}_l^{HH}$) from the values of a transformed image band. Hence, data compression ratio may depend on the order of bands processing. Some experimental results confirming this point are shown in Fig. 7. As we can see, the compression ratio at the most "successful" (the "best") order of bands processing is considerably (by 10-15%) higher than the compression ratio obtained for the same data at the "worst" order of bands processing.

One of the methods to define the most "successful" (suitable) combination of bands processing can be the enumeration of all possible combinations characterized by the need to carry out $K!$ compression operations with initial image and hence, by considerable computational expenses. To reduce computational expenses while defining the most "successful" order of bands processing it is suggested to use a method of "trimmed" bands enumeration. The method includes sequential increase in the number of bands wherein combination of 2 elements taken K at a time, then combination of 3 elements taken K at a time, etc. is subject to maximum compression ratio.

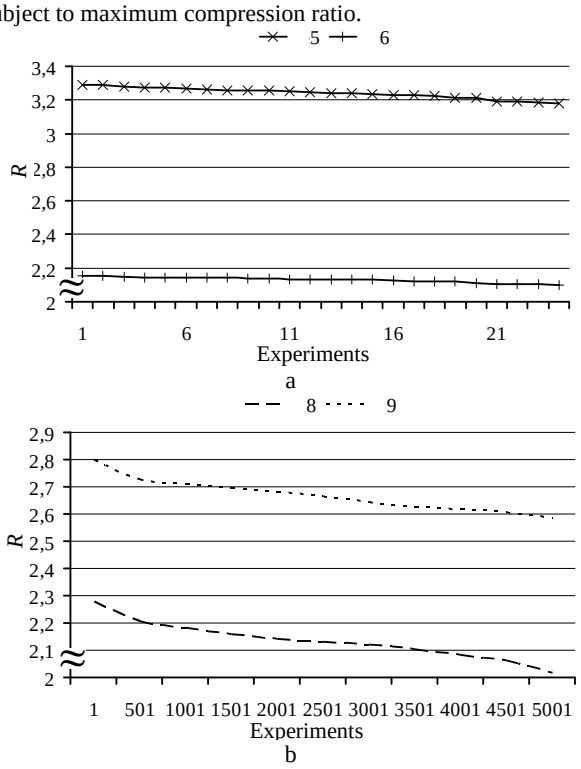


Figure 7. Compression ratio of multispectral RS images vs. different orders of bands enumeration and processing: $K\geq3$ (*a*), $K\geq7$ (*b*).

The algorithm of such trimmed enumeration given that vector with the length K, each element of which presents the number of a corresponding band processing in a corresponding position and can be shown in the following way:

Step 1. Set initial length of sub vectors v_i as $S=2$, $i=1,2,\ldots,C_S^K$, where C_S^K – the number of combinations of K elements taken S at a time.

Step 2. If $S = 2$ is formed from the initial vector $\mathbf{V}$ all possible sub vectors v_i with numbers of S length bands processing order.

Step 3. If $S > 2$ is formed from sub vector $\mathbf{v}_i^{\max}$ and every remaining bands of $\mathbf{V}$ vector all possible S length sub vectors $\mathbf{v}_i$.

Step 4. For all i carry out compression using three-stage algorithm taking into account bands processing order set in sub vectors $\mathbf{v}_i$.

Step 5. Out of all $\mathbf{v}_i$ find sub vector $\mathbf{v}_i^{\max}$ with maximum compression ratio.

Step 6. If $S < K–1$, then $S=S+1$, step 3, otherwise step 7.

Step 7. End.

K length sub vector v_i containing obtained (the "best") order of bands processing will be the result of the algorithm.

It is evident that the "best" order of bands processing obtained by this method can be different from the one obtained via complete enumeration of possibilities but at the same time it requires the less number of operations. In order to prove this point of view and to assess the compression ratio *R* of different test RS images a number of experiments was carried out, the results of which are shown in Fig. 6.

It is noteworthy that the RS image no.10 was excluded from the experiments due to practical impossibility to process 12 bands simultaneously. Obtained results demonstrate the advantage of trimmed enumeration in defining the processing order before any other method from the point of view of computational performance and compression ratio (in some cases slightly yielding to the maximum possible compression ratio obtained when using the completed enumeration of processing possibilities).

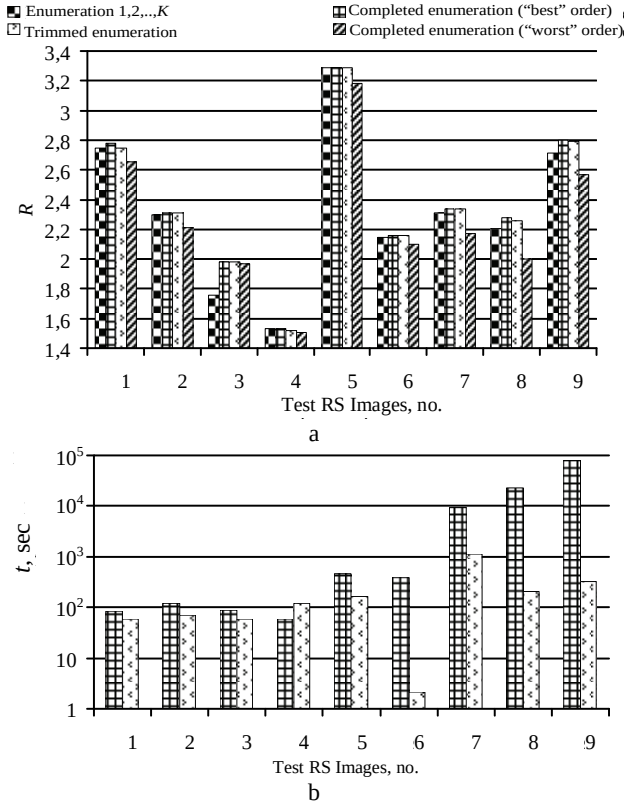


Figure 8. Relation of compression ratio and time on the way of bands processing order when identifying the "best" combination

Besides, when the number of image bands *K*>4 the advantage of the suggested method of trimmed enumeration becomes multiple, particularly for RS images of larger geometrical size (image no.6, Fig.*b*). The results presented in Fig. *b* show that even the use of trimmed enumeration might require dozens of seconds during compression, which cannot be considered satisfactory for practical purposes.

In order to increase the computational performance of defining the "best" order of bands processing it is suggested to modify the above considered trimmed enumeration algorithm by the selective data processing where step 4 should be applied not for the whole initial image but it is advisable to use randomly selected data samples of a specified size.

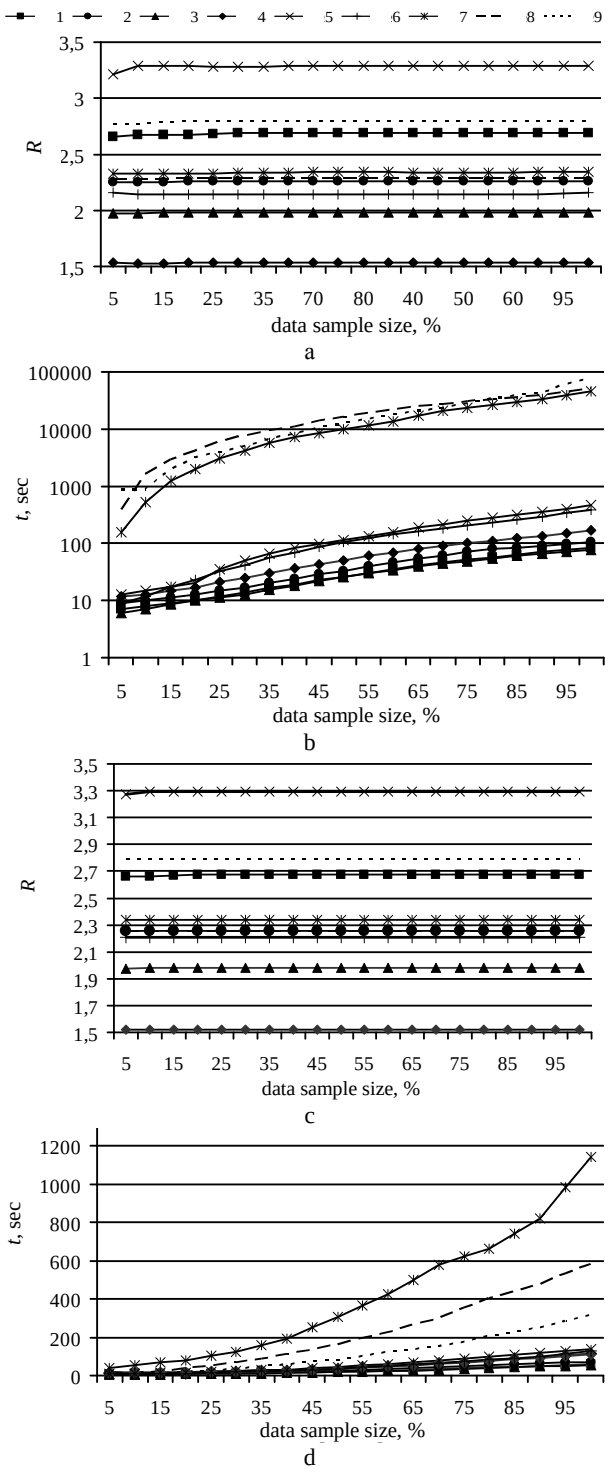


Figure 9. Relation of compression ratio and computational performance vs. data set size at completed (*a*,*b*) and trimmed (*c*,*d*) enumeration of bands

The results justify greater computational performance of the trimmed enumeration with the selective data use in compari Comparative investigation results of the prototypes and the modified algorithm based on methods of the trimmed enumeration, processing of bands and selective data processing with different data sample size are shown in Fig. 9.

son with the completed enumeration, as well as research allows one to make a conclusion on the fact that in order to define the "best" bands order processing it is enough to use lossless compression with a data sample size containing 8-10% of pixels of

initial data. In this case multiple increase of the computational efficiency (7÷40 times) is achieved. Thus, higher computational performance are typical for the images with more bands and greater geometrical size of scene. Based on experimental results and conclusions let us accept the use of selection equal 10% sufficient for the suggested modified algorithm.

The peculiar feature of the suggested compression algorithm is a joint use of wavelet transformations taking into account intra-bands correlation and some original modifications on search of the suitable order of the bands processing. The algorithm might increase possible compression ratio and improve computational performance. This allows one to advance universal *WinRar*, *WinZip* or *JPEG2000*, which do not consider peculiarities of multispectral RS images. However, the results of carried out comparative experiments presented in Fig. 10 show that suggested modifications of the compression algorithm and the selective data use with a various sample data size allow one to make the computational performance of the three-stage algorithm slightly comparable with *WinRar*, *WinZip* and *JPEG2000*, but at the same time more competitive in terms of compression ratio.

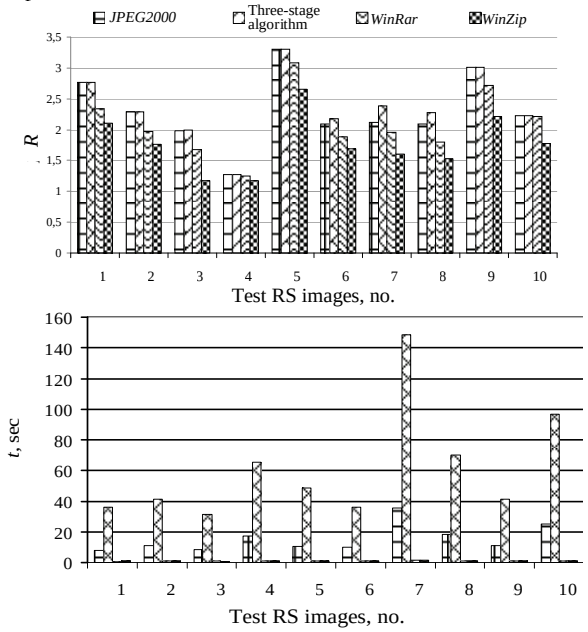


Figure 10. Comparative efficiency of the different compression algorithms

4. CONCLUSIONS

The three-stage compression algorithm of multispectral RS images based on the use of wavelet transformations and intra-bands correlation was developed. Data test set included 10 multispectral RS images of different RS systems was used for its research on a computational performance, compression ratio and validity. It allows one to define the ideal depth of wavelet transformation L=3, as well as to identify the relation of a compression ratio and time with the different ways of bands processing order while looking for the "best" bands combination from a view point of a maximum possible compression ratio. In order to increase the computational performance the algorithm was modified by a trimmed enumeration of the bands processing and data selective use, as well as the sufficient sample size equal 10% is identified. The overall computational performance increased in 7÷40 times, and compression ratio – on 15-20% in comparison with non-modified prototype.

The comparative research of the three-stage compression algorithm with universal *WinRar*, *WinZip* and *JPEG2000* was carried out taking into account all modifications and found parameters of the algorithm. Results allow one to make the conclusion on the considerable advantage of the suggested modified algorithm if compared it with its prototypes in terms of compression ratio but with some disadvantage in computational performance.

REFERENCE

Cagnazzo, M., Cicala, L., Poggi, G., Verdoliva, L., 2006. Low-complexity compression of multispectral images based on classified transform coding. *Signal Processing: Image Communication*. 10 (21), pp. 850-861.

Christopoulos, C., Skodras, A., Ebrahimi, T., 2001. The JPEG2000 still image coding system. *Signal Processing Magazine, IEEE*. 5(18), pp. 36-58.

Jacquin, A., 1993. Fractal Image Coding: A review. In: *Proceedings of the IEEE*, Vol. 81, pp 1451-1465.

Kiely, A., Klimesh, M., Xie, H., Aranki, N., 2006. ICER-3D: A Progressive Wavelet-Based Compressor for Hyperspectral Images. The Interplanetary Network Progress Report, pp. 142-164.

Marcellin, M., Abousleman, G., Hunt, B., 1995. Compression of hyperspectral imagery using the 3d dct and hybrid dpcm/dct. *J IEEE Transactions on Geoscience and Remote Sensing*, 33(1), pp. 26-35.

Motta, G., Rizzo, F., Storer J.A., 2006. *Hyperspectral Data Compression*. Springer, US, XI, 415 p.

Salmon, D., 2007. *Data Compression*. Springer-Verlag Ltd., London, p.1074

Taubman D., Marcellin M., 2002. *JPEG2000: Image Compression Fundamentals, Standard and Practice*. Kluwer Academic Publishers, Massachusetts, 808 p.

Witten, I., Neal, R., Cleary, J., 1987. Arithmetic coding for data compression. *J Communications of the ACM*. 30, pp. 520-540.

Ziv, J., Lempel, A., 1977. A universal algorithm for sequential data compression. *IEEE Transactions on Information Theory*. 23, pp. 337-343.

MONITORING IMPERVIOUS SURFACE SPRAWL USING TASSELED CAP TRANSFORMATION OF LANDSAT DATA

Qian Zhang [a, *], Yifang Ban [a]

[a] Geoinformatics Division, Dept. of Urban Planning and Environment,
Royal Institute of Technology-KTH, Stockholm, Sweden, 10044 - (qian.zhang, yifang.ban)@abe.kth.se

KEY WORDS: Urban, Monitoring, Change Detection, Transformation, Landsat

ABSTRACT:

Transformation from non-impervious surface to impervious surface changes the landscapes as well as the ecological and environmental conditions. Detecting impervious surface growth is vital to monitoring urban development and supporting sustainable city planning. The objective of this research is to conduct detection of impervious surface sprawl using tasseled cap transformation within the conceptual framework of Vegetation-Impervious surface-Soil (V-I-S) model. Landsat-3 MSS images on August 4, 1979 and Gap-filled Landsat-7 ETM+ images on May 22, 2009, covering the Greater Shanghai Area, were used in the case study. The results demonstrated that direct change detection using variables derived from tasseled cap transformation was effective for monitoring impervious surface sprawl. The variables derived from tasseled cap transformation have the potential to link to the components of the V-I-S model. The Greater Shanghai Area experienced high-speed impervious surface sprawl over the past 30 years at the average speed of 38.84km^2/year.

1. INTRODUCTION

The unprecedented combination of economic and population growth has led China into transition from a largely rural society to a predominantly urban one. Monitoring urban dynamic is of critical importance for urban planning and sustainable development in China, especially in the high-speed urbanization regions, such as the Greater Shanghai Area (GSA).

Remote sensing, as a "unique view" of the spatial and temporal dynamics of the processes of urban growth and land use change, has been widely used to monitor land cover changes. Many change detection methods and their improved versions have been investigated widely in the last two decades. Lu et al. (2004) summarized the change detection techniques into seven categories: algebra, transformation, classification, advanced models, GIS-related, visual analysis and other seldom-used techniques. Different approaches have their own advantages and disadvantages. It is impossible to say which approach is absolutely superior to the others, and sometimes different kinds of methods are combined so that the detection result is improved (Jensen, 2005; Seto et al., 2002).

Urban landscapes are a complex mix of buildings, roads, cloverleaf junctions, greenbelts, gardens, exposed soil and water body. In order to understand the heterogeneous landscapes better, the standardized model that describes these component surfaces should be defined. To date, the Vegetation-Impervious surface-Soil (V-I-S) model (Ridd, 1995) is one of the most commonly used conceptual models for remote sensing analysis of urban landscapes. The V-I-S model assumes that land cover in urban areas is a linear combination of three components: vegetation, impervious surface and soil (Figure 1). The diverse nature of these three substances has a significant impact on the dynamics and distribution of energy and moisture flux, the most important drivers in the ecosystem. It is a potentially powerful tool, therefore, for environmental impact analysis of urbanization such as urban heat island analysis (Weng & Lu, 2009). It is also serving as a global standard platform for spatial-temporal analysis and comparison of urban morphology, biophysical and human systems universally (Ridd, 1995).

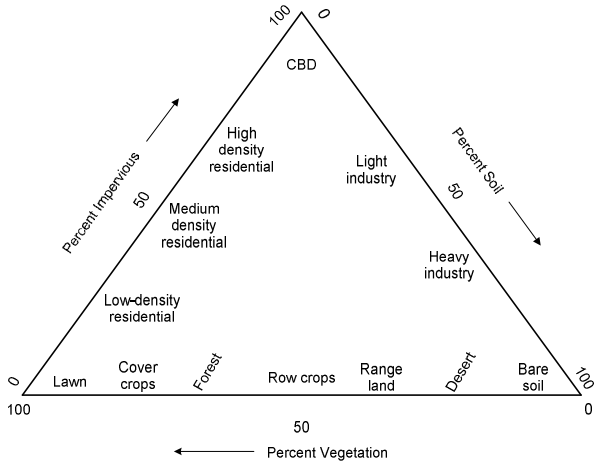


Figure 1. Some urban and near-urban features in the ternary V-I-S model (Ridd, 1995)

Many researchers have experimented with various methods to match the classifications of urban areas with the V-I-S conceptual model. Traditional unsupervised and supervised classification methods, for example, iterative self-organizing data analysis (ISODATA) and maximum likelihood classification (Madhavan et al, 2001); spectral unmixing analysis (Phinn et al., 2002; Weng & Lu, 2009); hierarchy classification (Setiawan et al, 2006; Ward et al., 2000) are the most frequently used methods. The original bands and the bands after principle components analysis (PCA) or minimum noise fraction (MNF) transformation of Landsat TM/ETM+ are the most frequently used data.

Tasseled cap transformation (so called K-T transformation), on other hand, highlights the characteristics of vegetation,and soil,

* Corresponding author

thus may have the potential to be related to the V-I-S components. According to Jensen (2005), urbanized areas are particularly distinct in the brightness component. The greater biomass covering, the brighter the pixel value in the greenness image. The wetness layer provides subtle information about the moisture status of the wetland environment. There has been very little research investigating whether tasseled cap transformation is effective for impervious surface sprawl detection and what is the relationship between variables derived from tasseled cap transformation and the components of the V-I-S model. Thus, there is a need to evaluate whether tasseled cap transformation is an appropriate tool for detecting impervious surface in the conceptual framework of the V-I-S model.

The objective of this research is to conduct detection of impervious surface sprawl in Shanghai using tasseled cap transformation within the conceptual framework of the V-I-S model. The specific aims of this research were: (1) to detect the impervious surface sprawl using tasseled cap transformation; (2) to analyze the spatial-temporal dynamics of impervious surface in GSA over the past 30 years; (3) to determine the relationship between the variables derived from tasseled cap transformation and the components of the V-I-S model.

2. STUDY AREA AND DATA DESCRIPTION

2.1 Study Area

GSA, located in the Yangtze River Delta, Eastern China, was selected as the case study area. It covers approximately 6,430 km^2 and owns the largest population (18.9 million persons in 2008) among all Chinese cities. Shanghai contributed 4.9% (approximately $129.5 billion dollars) of national Gross Domestic Product (GDP) in 2006, is one of the most affluent regions in China. During the past 30 years, Shanghai experienced a distinct urban sprawl owning to China's Reform and Open policy, as a result, major land cover transformations took place in Shanghai, where non-impervious surface (for example vegetation, water body, unused land) was replaced by the intensive build-up areas.

2.2 Data Description

To investigate impervious surface sprawl of Shanghai in the past 30 years, Landsat MSS images on August 4, 1979 and ETM+ images on May 22, 2009 were used (Figure 2). Past research has indicated that gap-filled Landsat-7 ETM+ SLC-off data with the USGS EROS released method is an accurate and acceptable data source and likely to produce acceptable classification products and to perform change detection (Bédard, et al., 2008).

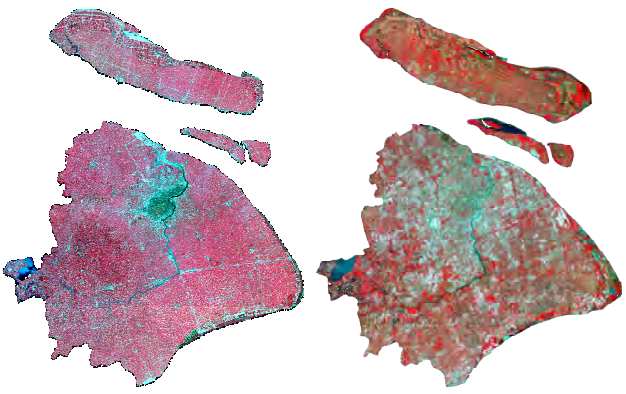


Figure2. Landsat images of Greater Shanghai Area, China in 1979 (Left, RGB: 321) and 2009 (Right, RGB: 432)

3. METHODOLOGY

3.1 Image Pre-processing

Landsat-7 ETM+ SLC-off data was gap-filled using segment-based method (USGS, 2004). Then the Landsat MSS and ETM+ images were geocoded with a root mean square error (RMSE) approximately 6 m. As a single scene could not cover the whole study area, image mosaicking of two adjacent scenes was carried out and color balance was performed in the overlap region between the two scenes. Then the image was subsetted to the boundary of the GSA.

3.2 Tasseled Cap Transformation (K-T Transformation)

Tasseled cap transformation was developed by Kauth and Thomas in 1976 for Landsat MSS data (Kauth & Thomas, 1976) and was improved and extended to Landsat TM data in the mid-1980s (Crist & Cicone, 1984a; Crist, 1985; Crist & Kauth, 1986). It was widely used in monitoring agriculture, vegetation changes (Collins & Woodcock, 1996; Han et al., 2007; Price et al., 2002; Rogan et al., 2002) as well as urban dynamics detection (Fung, 1990; Seto et al., 2002). Tasseled cap transformation is a kind of orthogonal transformation, it rotates the original data plane so that the vast majority of data variability is concentrated in the features, i.e., the plane is viewed "head-on" and it presents the most basic structures of the data in the most direct way (Crist & Cicone, 1984b). The original Landsat MSS data space was transformed to a new four-dimensional feature space, that is, the soil brightness index (B), greenness vegetation index (G), yellow stuff index (Y), and non-such (N). In this study, we aimed to extract built-up areas information and B, G and Y were derived.

$$Brightness = 0.332M_1 + 0.603M_2 + 0.675M_3 + 0.262M_4 \quad (1)$$

$$Greenness = -0.283M_1 - 0.660M_2 + 0.557M_3 + 0.388M_4 \quad (2)$$

$$Yellowness = -0.899M_1 + 0.428M_2 + 0.076M_3 - 0.041M_4 \quad (3)$$

Where $M_{1\text{-}4}$ represent digital numbers (DNs) of band 1 to 4.

For Landsat-7 ETM+ images, Brightness (B), Greenness (G), Wetness (W) and the fourth, fifth, sixth variables were produced and the first three variables were selected. Past research has indicated that at-satellite reflectance-based K-T transformation is needed because changing sun illumination geometry affects DNs strongly, and thus affects the derived tasseled cap value. Moreover, at-satellite reflectance-based K-T transformation can differentiate water from land targets better than the DNs and reflectance factor based transformation (Huang et al., 2005). DNs should, therefore, be converted to at-satellite reflectance firstly according to formulae 3 and 4 (Myeong et al., 2006). After the two steps, the effects owing to different sun angles at different dates were compensated. Correspondingly, the coefficients (Huang et al., 2005) here varied from the DNs-based transformation (formulae 5-7).

$$L_\lambda = (DN_\lambda \cdot gain_\lambda) + bias_\lambda \quad (3)$$

Where L is radiance; λ is the spectral band; gain is he spectral band gain; and bias is the spectral band offset.

$$\rho_\lambda = \frac{\pi \cdot L_\lambda \cdot d^2}{E_{sun\lambda} \cdot \cos(\theta)} \quad (4)$$

Where ρ is the unitless in-band planetary reflectance; λ is the spectral band; L is radiance; d is the Earth-Sun distance; E_{sun} is the mean solar atmospheric irradiance; and θ is solar zenith angle in degrees.

$$Brightness = 0.3561\,\rho_1 + 0.3972\,\rho_2 + 0.3904\,\rho_3 + 0.6966\,\rho_4 + 0.2286\,\rho_5 + 0.1596\,\rho_7 \quad (5)$$

$$Greeness = -0.3344\,\rho_1 - 0.3544\,\rho_2 - 0.4556\,\rho_3 + 0.6966\,\rho_4 - 0.0242\,\rho_5 - 0.2630\,\rho_7 \quad (6)$$

$$Wetness = 0.2626\,\rho_1 + 0.2141\,\rho_2 + 0.0926\,\rho_3 + 0.0656\,\rho_4 - 0.7629\,\rho_5 - 0.5388\,\rho_7 \quad (7)$$

3.3 Detection of Impervious Surface Sprawl

3.3.1 Direct Change Detection

Direct change detections were performed using Brightness and Greenness variables from 1979 and 2009. As the transition from non-impervious surface to impervious surface results in big difference in both brightness and greenness, the sprawl area should be concentrated in upper-right part of the 2D scatter plot between Δ Brightness and Δ Greenness (Figure 3a). By adjusting the data range (points in red region of figure 4a) in the plot, the corresponding distribution of points can be highlighted interactively in the image window (Figure 3b). The impervious surface sprawl was then determined with the assistance of field data and the Landsat images.

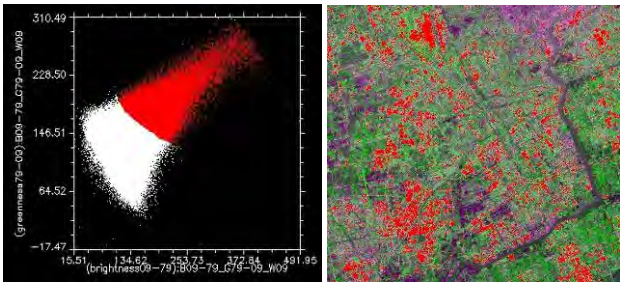


Figure 3(a) 2D scatter plot between ΔBrightness (X-axis) and ΔGreenness (Y-axis) (b) Corresponding regions to red points of 2D scatter plot of Figure 4a

3.3.2 Unsupervised Classification: ISODATA

Unsupervised classification was performed on the changes in Brightness and Greenness using Interactive Self-Organizing Data Analysis (ISODATA). The parameters for ISODATA classification conclude: change threshold is 5%, maximum class stdv is 1.00, minimum class distance is 5.00 and the number of classes is from 5 to 10.

3.4 Accuracy Assessment

For validation of the change detection results, 2000 randomly selected testing points were generated, among which 1000 were for impervious surface sprawl (ISS) and the other 1000 for non-impervious surface (NISS).

4. RESULTS AND DISCUSSION

4.1 Visualization of the Variables from K-T Tramsform

Figure 4 shows the colour composites of Brightness, Greenness, and Wetness derived from the 2009 ETM+ images. Built-up areas are highlighted in Red and vegetation is in Green while water is in Blue.

Figure 4. Colour composites map of Brightness, Greenness,and Wetness derived from 2009 ETM+ data

Figure 5 shows the colour composites of Brightness, Greenness, and Yellowness derived from the 1979 MSS images. Built-up areas are shown in blue and orange, vegetation is in Yellow while water is in dark blue, purple and black.

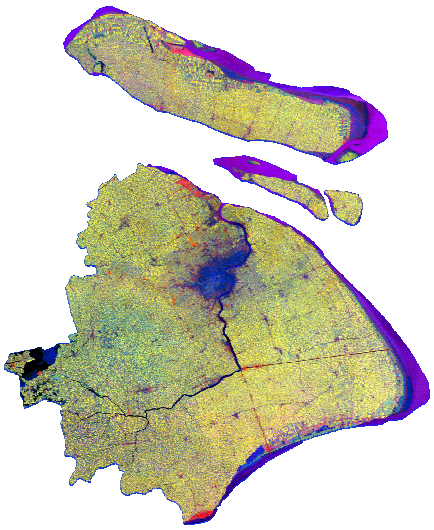


Figure 5. Colour composites map of Brightness,Greenness,and Yellowness derived from 1979 MSS data

Figure 6 is the colour composites of changes in Brightness (in Red and Blue) and Greenness (Green) in 2009. ΔBrightness is derived by brightness 2009 minus brightness 1979 while ΔGreenness is computed by greenness 1979 minus greenness 2009 in order to highlight the impervious surface sprawl. As a result, ΔBrightness above 0 means brightness increases while ΔGreenness above 0 means greenness decreases. Increase in brightness and decrease in greenness is highlighted in white in Figure 6, and the single change in brightness or greenness is highlighted in purple or green correspondingly.

From the observation of Figure 6 with the aid of field investigations, most impervious surface sprawl is related to

white regions except the coastal area. It indicated that, on one side, the impervious surface sprawl can be highlighted well in the color composites of increases in Brightness and decreases in Greenness. Even though the images were acquired in late May and early August respectively, seasonal effects were not significant as both late May and Early August belong to the summer season with the multi-year average temperature in May at 22°C and in August at 28.7°C and vegetation in Shanghai in both months is green. On the other hand, impervious surface sprawl is not characterized in white in shoreline areas. This maybe explained according to Jensen (2005), when conducting change detection in the coastal zone, many influence factors should be considered, such as the different tidal stage.

Figure 6. Color composites of ΔBrightness (R), ΔGreenness (G) and ΔBrightness (B)

4.2 Detection of Impervious Surface Sprawl

The direct change detection results using 2-D scatter plot of Brightness and Greenness is shown in Figure 7, with the increase in impervious surface in yellow, and the background map is Figure 6. The accuracy assessment is listed in Table 8. Greater than 90% of the producers' and users' accuracies were achieved and the kappa is 0.84.

Detected sprawl area of impervious surface

Figure 7. Detected result of impervious surface sprawl using 2D scatter plot

Ref. / Res.	ISS	NISS	Total	Producers' Accu.	Users' Accu.
ISS	901	82	983	90.1%	91.7%
NISS	99	918	1017	91.8%	90.3%
Total	1000	1000	2000	91.0%	

Kappa Coefficient= 0.84

Table 8. Accuracy of direct change detection

For the unsupervised classification, seven classes were produced and the 7th class was determined as impervious surface sprawl while the other six classes were recognized as non-impervious surface. The impervious surface sprawl result is presented in yellow in Figure 9. Figure 6 is taken as the background map too. The accuracies were lower than that of the direct change detection (Table 10).

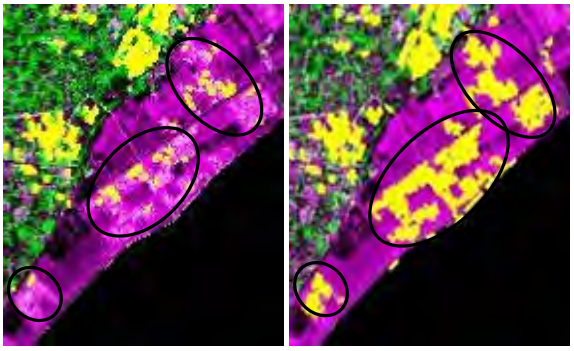
Detected sprawl area of impervious surface

Figure 9. Detected result of impervious surface sprawl using ISODATA unsupervised classification

Ref. / Res.	ISS	NISS	Total	Producers' Accu.	Users' Accu.
ISS	851	85	936	85.1%	90.9%
NISS	149	915	1064	91.5%	86.0
Total	1000	1000	2000	88.3%	

Kappa Coefficient= 0.80

Table 10. Accuracy of ISODATA unsupervised classification



Detected sprawl area of impervious surface

Figure 11. Comparison of diction method in shoreline area (a) direct change detection (b) ISODATA classification

Figure 11 shows the comparisons of the two methods in the shoreline regions. Detected impervious surface sprawl was highlighted in yellow. The difference in the ellipse illustrated that ISODATA classification was able to detect nearly all the impervious surface sprawl while the large protion of the impervious surface sprawl was not detected using direct change detection. Visual inspection indicated that in this case ISODATA is superior for detecting impervious surface in the shoreline region even though direct change detection method has the higher overall accuracy than ISODATA. Therefore, the final map of the impervious surface sprawl was a composite of both results with ISODATA result in the shoreline area and direct change detection result in other regions.

The overall accuracy of combined detection method on both ISS and NISS is 92.5%, and the Kappa coefficient is 0.8614. The detected impervious surface sprawl result is shown in Figure 13 and the detailed accuracy is listed in Table 12.

Ref. / Res.	ISS	NISS	Total	Producers' Accu.	Users' Accu.
ISS	913	63	976	91.3%	93.5%
NISS	87	937	1024	93.7%	91.5%
Total	1000	1000	2000	92.5%	

Kappa Coefficient= 0.86

Table 12. Accuracy of hybrid detection method

4.3 The spatial-temporal dynamics of impervious surface in Shanghai

Based on combined result, the temporal-spatial dynamics of impervious surface in GSA over the past 30 years was analyzed. The impervious surface encroached large area of agricultural land and other vegetation (approximate 1165.1km^2) over the past 30 years with the extremely high annual expansion speed (38.84 km^2/year in average). In addition, Shanghai reclaimed and is reclaiming land from the sea due to many pressures, such as the high land price in the city, and agriculture compensation policy for build-up encroached areas.

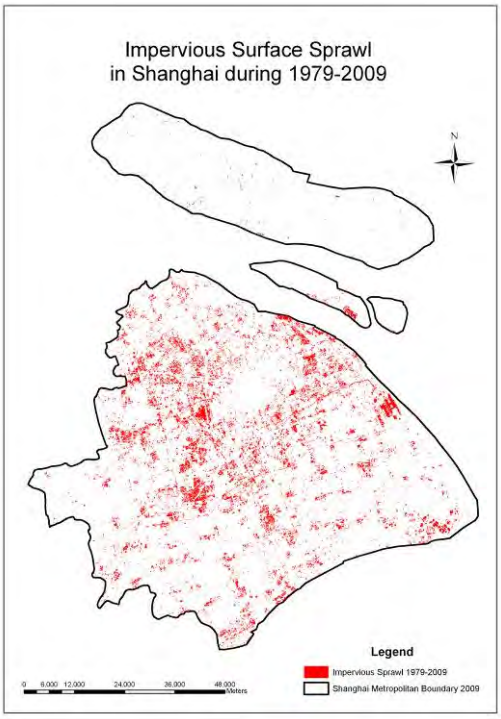


Figure 13. Detection of impervious surface sprawl between 1979 & 2009

4.4 The relationship between Tasseled Cap Variables and the V-I-S Components

In the following 2D scatter plots (Figure 14 a, b), impervious surface, soil and low coverage of vegetation, water body and high coverage of vegetation is expressed in red, yellow, blue and green respectively. The samples are selected using the ground truth points from field work in October, 2009. From the following figures, the different components of the V-I-S model can be illustrated relatively separated in the feature space of tasseled cap variables, for instance Brightness-Greenness space, than the feature space of Landsat ETM+ original bands.

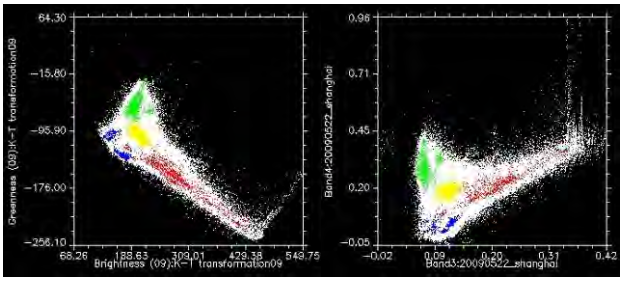


Figure 14 (a). Scatter plot between Brightness (X-axis) and Greenness (Y-axis) in 2009; (b). Scatter plot between Band 3 (X-axis) and Band 4 (Y-axis) of Landsat ETM+, 2009

5. CONCLUSION

This study showed that tasseled cap transformation is an effective method for detecting impervious surface sprawl in GSA and the variables derived from tasseled cap transformation has the potential to link with the components of the V-I-S model. The results also demonstrated that the GSA experienced high-speed impervious surface sprawl over the past 30 years at the average speed of 38.84km^2/year.

REFERENCES

Bédard, F., Reichert, G., Dobbins, R., & Trépanier, I., 2008. Evaluation of segment-based gap-filled Landsat ETM+ SLC-off satellite data for land cover classification in southern Saskatchewan, Canada. International Journal of Remote Sensing, 29(7), 2041-2054.

Collins, J. B., & Woodcock, C. E., 1996. An assessment of several linear change detection techniques for mapping forest mortality using multitemporal landsat TM data. Remote Sensing of Environment, 56(1), 66-77.

Crist, E. P., & Cicone, R., 1984a. Application of the tasseled cap concept to simulated Thematic Mapper data. Photogrammetric Engineering and Remote Sensing, 50(3), 343-352.

Crist, E. P., & Cicone, R. C., 1984b. A physically-based transformation of thematic mapper data---The TM Tasseled Cap. Geoscience and Remote Sensing, IEEE Transactions on, GE-22(3), 256-263.

Crist, E. P., & Kauth, R. J., 1986. The Tasseled Cap De-Mystified. Photogrammetric Engineering and Remote Sensing, 52(1), 81-86.

Crist, E. P., 1985. A TM Tasseled Cap equivalent transformation for reflectance factor data. Remote Sensing of Environment, 17(3), 301-306.

Fung, T., 1990. An assessment of TM imagery for land-cover change detection. IEEE Transactions on Geoscience and Remote Sensing, 28(4), 681-684.

Han, T., Wulder, M., White, J., Coops, N., Alvarez, M., & Butson, C. 2007. An Efficient Protocol to Process Landsat Images for Change Detection With Tasselled Cap Transformation. Geoscience and Remote Sensing Letters, IEEE, 4(1), 147-151.

Huang, C., Wylie, B., Zhang, L., Homer, C., & Zylstra, G., 2005. tasseled.pdf (application/pdf Object). Retrieved September 11, 2009, from http://landcover.usgs.gov/pdf/tasseled.pdf

Jensen, J. R., 2005. Introductory Digital Image Processing A Remote Sensing Perspective (3rd ed.). Prentice Hall.

Kauth, R., & Thomas, G., 1976. The tasselled cap-a graphic description of the spectral-temporal development of agricultural crops as seen by LANDSAT. In Proceedings of the Symposium on Machine Processing of Remotely Sensed Data (pp. 4B41-4B51). Presented at the Proceedings of the Symposium on Machine Processing of Remotely Sensed Data, Purdue University, West Lafayette.

Lu, D., Mausel, P., Brondízio, E., & Moran, E., 2004. Change detection techniques. International Journal of Remote Sensing, 25(12), 2365-2401.

Madhavan, B. B., Kubo, S., Kurisaki, N., & Sivakumar, T. V. L. N., 2001. Appraising the anatomy and spatial growth of the Bangkok Metropolitan area using a vegetation-impervious-soil model through remote sensing. International Journal of Remote Sensing, 22(5), 789-806.

Myeong, S., Nowak, D. J., & Duggin, M. J., 2006. A temporal analysis of urban forest carbon storage using remote sensing. Remote Sensing of Environment, 101(2), 277-282.

Phinn, S., Stanford, M., Scarth, P., Murray, A. T., & Shyy, P. T., 2002. Monitoring the composition of urban environments based on the vegetation-impervious surface-soil (VIS) model by subpixel analysis techniques. International Journal of Remote Sensing, 23(20), 4131-4153.

Price, K. P., Guo, X., & Stiles, J. M., 2002. Optimal Landsat TM - band combinations and vegetation indices for discrimination of six grassland types in eastern Kansas. International Journal of Remote Sensing, 23(23), 5031-5042.

Ridd, M. K., 1995. Exploring a V-I-S (vegetation-impervious surface-soil) model for urban ecosystem analysis through remote sensing: comparative anatomy for cities†. International Journal of Remote Sensing, 16(12), 2165-2185.

Rogan, J., Franklin, J., & Roberts, D. A., 2002. A comparison of methods for monitoring multitemporal vegetation change using Thematic Mapper imagery. Remote Sensing of Environment, 80(1), 143-156.

Setiawan, H., Mathieu, R., & Thompson-Fawcett, M. 2006. Assessing the applicability of the V-I-S model to map urban land use in the developing world: Case study of Yogyakarta, Indonesia. Computers, Environment and Urban Systems, 30(4), 503-522. doi:10.1016/j.compenvurbsys.2005.04.003

Seto, K. C., Woodcock, C. E., Song, C., Huang, X., Lu, J., & Kaufmann, R. K. 2002. Monitoring land-use change in the Pearl River Delta using Landsat TM. International Journal of Remote Sensing, 23(10), 1985-2004.

USGS. 2004. SLC-off Gap-Filled Products Gap-Fill Algorithm Methodology. Retrieved September 11, 2009, from http://landsat.usgs.gov/documents/L7SLCGapFilledMethod.pdf

Ward, D., Phinn, S. R., & Murray, A. T. 2000. Monitoring growth in rapidly urbanizing areas using remotely sensed data. Professional Geographer, 52(3), 371-386.

Weng, Q., & Lu, D. 2009. Landscape as a continuum: an examination of the urban landscape structures and dynamics of Indianapolis City, 1991–2000, by using satellite images. International Journal of Remote Sensing, 30(10), 2547-2577.

ACKNOWLEDGEMENT

This research was also supported by grants from the Swedish Science Foundation (VR), the Swedish Science Foundation and the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS) awarded to Professor Ban.

PRECISE PROCESSING OF SPOT-5 HRS AND IRS-P5 STEREO IMAGERY – FOR THE PROJECT OF WEST CHINA TOPOGRAPHIC MAPPING AT 1:50,000 SCALE

L. Zhang * , J. Zhang

Chinese Academy of Surveying and Mapping, Haidian, Beijing, 100039 - {zhangl,zhangjx}@casm.ac.cn

KEYWORDS: High-Resolution, Model, Adjustment, Image, Matching, DTM

ABSTRACT:

High-resolution satellite images (HRSI) at sub-5m footprint such as IKONOS, IRS-P5 (CartoSat-1) and SPOT-5 HRG/HRS images are the main data sources for the Project of West China Topographic Mapping (WChTM), which has been approved by the State Council of China in 2006 and will be completed at the end of 2010. Two procedures which are used for this project in practice, i.e. the block-adjustment procedure and automatic DTM generation procedure, are described in this paper. We firstly present an approach for block-adjustment based on Rational Function Model (RFM) with sparse GCPs by using satellite Images. Secondly, we present a matching approach for automatic DTM generation from HRSI.
To test the proposed approaches, they have been applied to SPOT-5 images over 1 test-fields, which covers eastern part of Tibet Plateau, China with variable terrain geomorphologic type. In another test we use 23 scenes of IRS-P5 images, which cover Beijing test area of about 21,000 square kilometers. From these experiments, it's shown that with the proposed block-adjustment and DTM generation approach, by using SPOT-5 HRS/HRG and IRS-P5 imagery with small number of GCPs, satisfactory image orientation results and DTM product (after necessary manual editing) can be achieved with a little bit better accuracy than those requirements from the Chinese Surveying and Mapping regulations for 1:50000 topographic maps.

1. INTRODUCTION

A decade after early 1990s, remote sensing has stepped into a new stage which can supply various high-resolution observation data from space. At present, terrain information extraction, change detection, disaster monitoring, and topographic mapping by using the HRSI has become one of research hotspots. Meanwhile, HRSI has more and more applications in photogrammetry. The stereo remote sensing image with spatial resolution of meter-level or even sub-meter level has the capability to replace the aerial images which used for traditional topographic mapping or updating of geo-information at 1:50,000 and 1:10,000 scale (Zhang, et. al., 2004). Among these, GeoEye-1, IKONOS, SPOT-5 HRS/HRG, IRS-P5 images are well-known examples.

Due to the difficulties for aerial image acquisition, time limit of the project and other practical considerations, SPOT-5 HRS/HRG, IRS-P5 and other HRSI images are the main data sources for the project of Western China Topographic Mapping (WChTP) at 1:50,000 scale. Before 2006, up to 2.02 million km^2 are not mapped at 1:50,000 scale in western part of China, it includes Sorthern-XinJiang desert area, Qing-Tibet Plateau area and Heng-duan mountain ranges. This unmapped area covers about 20% of all areas of China; includes total number of about 5,032 sheets of 1:50,000 scale topographic maps. This situation greatly hiders the socio-economic development of this region, it also poses potential threat to national security. The project has been approved by the State Council of China in year 2006, and is dedicated to complete 1:50,000 scale topographic map and construct the national geo-spatial database for the region within next 5 years through year 2006 to 2010. The unmapped area

covers most unmans area of China, the very harsh natural conditions (average terrain elevation is more than 4000m) and difficult transportation conditions provide a very difficult situation for GCP measurement and other field works. In many areas, it's even impossible for accessing or collecting enough GCPs, therefore, the project of WChTP should utilize some innovative and well-developed techniques, such as DGPS/IMU assisted aerial photogrammetry, high resolution satellite imagery mapping technique and radar image mapping technique, design and use mapping procedure with small number of GCPs, to complete the project. However, on one hand, the radar image processing technique, especially those methods for precise DTM generation from InSAR data, has not been well-developed and regularly used in China; on the other hand, very high resolution aerial images cannot be acquired within reasonable time due to the very harsh natural conditions and variable climate conditions, Therefore, in the project of WChTP, more than 90% areas will be mapped with SPOT-5 HRS/HRG and IRS-P5 imagery (including IKONOS and QuickBird images over some important cities and towns), the rest of areas will be mapped with SAR and aerial images only after enough research works and experiments.

The project of WChTP is a complicated project which first uses HRSI for topographic mapping in China In this paper, we firstly present a block-adjustment procedure based on Rational Function Model (RFM) for HRSI satellite images (Chapter 2). Secondly, we present a matching approach for automatic DTM generation. It can provide dense, precise and reliable results (Chapter 3). These two procedures have been successfully applied for block-adjustment of large-area SPOT-5 and IRS-P5 satellite imagery with small number of GCPs, and used in practice in production-line for the project of WChTP. The results show that with our approaches, by using SPOT-5 HRS/HRG and IRS-P5 imagery, with several GCPs, satisfactory image orientation results and DTM products (after necessary manual editing) can be completed with a little bit better accuracy than those requirements from Chinese Surveying and Mapping regulations for 1:50000 topographic maps (Chapter 4, 5).

* Corresponding author

2. BLOCK-ADJUSTMENT WITH HRSI BASED ON RATIONAL FUNCTION MODEL (RFM) AND SPARSE NUMBER OF GCPS

Sensor models are fundamental for the photogrammetric processing, such as the stereo measurements and the image ortho-rectification. They are typically classified into two categories: the physical and the generalized models. In a generalized sensor model, the transformation between the image and the object space is represented as some general function without modeling the physical imaging process. The Rational Function Models (RFMs) is one of the generalized sensor models and have recently drawn considerable interest in the remote sensing community.

Almost all the high-resolution satellite cameras use Linear Array CCDs to acquire a single image line at an instant of time, each with its own positional and attitude data. The imaging geometry is characterized by nearly parallel projection in along-track direction and perspective projection in cross-track direction. The bundle adjustment approach, which has been well-developed in aerial photogrammetry, can also be applied for satellite images after appropriate alteration (Qian, et. al., 1990). Due to the dynamic nature of satellite image acquisition, this kind of model is more complicated than in the single frame case. Furthermore, due to very narrow field of view for HRSI images (e.g. SPOT-5 HRS is 8.3°, IKONOS only is 0.7°), many parameters in the physical sensor models are completely or highly correlated with other parameters so that they cannot be safely estimated through the triangulation procedure (Grodecki and Dial, 2003). According to the researches made by Tao and Hu (2001), The RFM can achieve an approximation accuracy that is extremely high both for aerial frame data and SPOT linear array data. Therefore, in this paper, we try to develop a block-adjustment approach with HRSI based on RFM. The procedure includes the following 2 steps:

2.1. Rational Function Model (RFM) parameter estimation

A RFM is generally the ratio of two polynomials with its parameters derived from the physical sensor model and the corresponding terrain information. In RFM, image pixel coordinates (x, y) are expressed as the ratios of polynomials of object coordinates (φ, λ, h), which in the case of the IKONOS RPCs correspond to latitude, longitude and ellipsoidal height. For an image, where x_n and y_n are normalised pixel coordinates and φ_n,λ_n,h_n are normalised latitude, longitude and ellipsoidal height, the ratios of polynomials have the following form:

$$x_n = RPC_x(\varphi,\lambda,h) = \frac{f_1(\varphi_n,\lambda_n,h_n)}{f_2(\varphi_n,\lambda_n,h_n)}$$
$$y_n = RPC_y(\varphi,\lambda,h) = \frac{f_3(\varphi_n,\lambda_n,h_n)}{f_4(\varphi_n,\lambda_n,h_n)} \quad (1)$$

In equation (1), the maximum power of each object coordinate and the total power of all object coordinates are limited to 3.

Some commercial HRSI like IKONOS and IRS-P5, only supply RFM model coefficients to the user, however SPOT-5 HRS/HRG supplies orientation parameters through the metadata file (DIMAP format file).

Table 1: Accuracy test for SPOT-5 imagery RPC Estimation

Image type	Image size (Row×Column)	Fitting RSME (pixel)
HRS 5×10m Pan	50712×12000	0.00854
HRS 5×10m Pan	116264×12000	0.01017
HRS 5×10m Pan	114592×12000	0.01016
HRS 5×10m Pan	81928×12000	0.00978
HRS 5×10m Pan	116864×12000	0.01013
HRG 5m Pan	40208×12000	0.02557
HRG 5m Pan	51984×12000	0.02814
HRG 2.5m Pan	24000×24000	0.05495
HRG 2.5m Pan	24000×24000	0.05518
HRG 10m MS	6000×6000	0.01099
HRG 10m MS	6000×6000	0.01558

Through these parameters, the physical sensor model (details please refer to SPOTIMAGING, 2002) of SPOT-5 imagery can be established. Usually, the RFM can be computed based on the SPOT-5 physical sensor model. Tao and Hu (2001) gave a detailed description of a least squares solution of RPCs based on the physical sensor models and suggested using a Tikhonov regularization for tackling possible oscillations. In our procedure, this method was applied also for SPOT-5 image RPC estimation.

We applied comprehensive testing by using hundreds scene of SPOT-5 HRS/HRG images in order to evaluate the performance of RPC estimation method, part of the results are shown in Table 1. The results show that: a) For SPOT-5 HRS stereo image, fitting RMSE for physical sensor model usually is about 1/100 pixel, and it seems not related to the image size; for SPOT-5 HRS 5m panchromatic images, fitting accuracy is about 1/40 pixel, for SPOT-5 HRS 2.5m panchromatic images, fitting accuracy is about 1/20 pixel. b) The fitting RMSE for physical sensor model is roughly related to ground resolution of the images, which is roughly equal to 0.1m in ground. c) RFM can replace the physical sensor model in subsequent photogrammetric processing of SPOT-5 image, and the accuracy loss can be ignored in mapping at 1:50,000 scale.

2.2. Block-adjustment based on RFM

Actually, the RFM constitutes a re-parameterization of the physical sensor model. Errors in sensor interior and exterior orientation thus give rise to errors in the RPCs. Grodecki and Dial (2003) proposed a practical block-adjustment model for multi-strip blocks of the high-resolution satellite imagery described by RPC models and illustrated the method with an IKONOS example. With the supplied RPCs, the mathematical model used is:

$$x_k + a_{i,0} + a_{i,1}x_k + a_{i,2}y_k = RPC_x^i(\varphi_k,\lambda_k,h_k)$$
$$y_k + b_{i,0} + b_{i,1}x_k + b_{i,2}y_k = RPC_y^i(\varphi_k,\lambda_k,h_k) \quad (2)$$

where, $a_{i,0}$, $a_{i,1}$, $a_{i,2}$ and $b_{i,0}$, $b_{i,1}$, $b_{i,2}$ are the 6 adjusted parameters for image i, and (x_k, y_k) and (φ_k, λ_k, h_k) are pixel and object coordinates of the points k.

In our approach, we first used the RPCs to transform from object to image space and then using these values and the known pixel coordinates we estimated either two shifts $a_{i,0}$, $b_{i,0}$ (model *M_RPC2*) or all 6 parameters $a_{i,0}$, $a_{i,1}$, $a_{i,2}$ and $b_{i,0}$, $b_{i,1}$, $b_{i,2}$ (model *M_RPC6*). The basic least squares observation equations for these 2 models are:

$$v = A\Delta + l; \quad P$$
$$v = [v_x, v_y]_k^T; \quad \Delta = [a_{1,0}, a_{1,1}, a_{1,2}, b_{1,0}, b_{1,1}, b_{1,2}, \Delta\varphi_k, \Delta\lambda_k, \Delta h_k]^T$$
$$\text{where} \quad A = \begin{bmatrix} 1, x_k, y_k, 0, 0, 0, \frac{\partial x_k}{\partial \varphi_k}, \frac{\partial x_k}{\partial \lambda_k}, \frac{\partial x_k}{\partial h_k} \\ 0, 0, 0, 1, x_k, y_k, \frac{\partial y_k}{\partial \varphi_k}, \frac{\partial y_k}{\partial \lambda_k}, \frac{\partial y_k}{\partial h_k} \end{bmatrix}_i; \quad l = \begin{bmatrix} x_k - RPC_x^i(\varphi_k, \lambda_k, h_k) \\ y_k - RPC_y^i(\varphi_k, \lambda_k, h_k) \end{bmatrix} \quad (3)$$

Here, P is the weight matrix describing the image measurement precision. For a block of images, the image georeferencing can be performed with the solution of the least squares normal equations resulting from equation (3). From equations (2) and (3) it is apparent that model *M_RPC2* requires at least a single well-defined GCP, whereas an estimation of all six parameters in model *M_RPC6* would require a minimum of 3 appropriately located GCPs per image.

3. AUTOMATIC DTM GENERATION BY USING A MULTIPLE-PRIMITIVE MULTIPLE-IMAGE MATCHING APPROACH

We have developed an advanced matching approach for automatic DTM generation from HRSI. The approach uses a coarse-to-fine hierarchical solution with a combination of several image matching algorithms and automatic quality control. The approach essentially consists of 3 mutually connected components: the image pre-processing, the multiple primitive multi-image (MPM) matching and the geomorphologic refinement matching procedure. The overall data flow is shown schematically in Fig. 1.

The Multiple Primitive Multi-Image (MPM) matching procedure is the core of our approach for accurate and robust DTM reconstruction. In this procedure, we do not aim at pure image-to-image matching. Instead we directly seek for image-to-object correspondences. We have developed a new flexible and robust matching algorithm – Geometrically Constrained Cross-Correlation (GC^3) method in order to take advantage of the multiple images. The algorithm is an extension of the standard Cross-Correlation technique and is based on the concept of multi-image matching guided from object space and allows reconstruction of 3D objects by matching all available images simultaneously, without having to match all individual stereo-pairs and merge the results. For more details of our matching procedure, please refer to Zhang, (2005); Baltsavias, et. al., (2006).

It should mention that the results through matching two or more optical remote sensing images are actually the Digital Surface Models, i.e. DSMs. Post-processing of DSMs are necessary to generate DTMs through methods which can reduce the DSM points to bare earth in urban areas or heavily vegetated areas. Manual editing the DSMs through stereo checking is one of the options; however, it is a time-consuming job. In our procedure, a method which used for LIDAR data filtering was used. The filtering method is one of the so-called morphological filters, which use a small structure element, describing admissible height differences as a function of the horizontal distance. This method was modified, implemented and integrated into our matching procedure as a post-processing option. For details of this filtering method, please refer to Vosselman, G., (2000).

4. BLOCK-ADJUSTMENT TESTING AND ACCURACY ANALYSIS

In order to evaluate the performance of block-adjustment approach presented in this paper. We selected several test areas to apply extensively accuracy test, the results show that only using small number of GCPs, we can meet the requirements of topographic mapping at 1:50,000 scale in China for large coverage of SPOT-5 HRS and IRS-P5 stereo images.

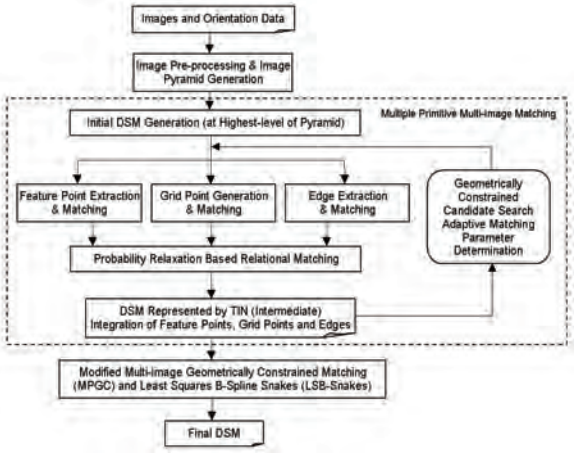


Figure 2. Workflow of the proposed automated DTM generation approach.

4.1 SPOT-5 HRS Imagery in Eastern Tibet Plateau and Eastern Talimu Basin

The block-adjustment with SPOT-5 HRS images in areas of eastern Tibet Plateau and eastern Talimu basin have been completed in 2006 and 2007. The test area in eastern Tibet Plateau covers 1234 topographic maps at 1:50,000 scale with the area of about 530,000km^2, where contains large-area of seasonally and perennially frozen soil, glacier and perennial snowfield and unman area; The test-field in eastern Talimu Basin and North Slope of Aerjin Mountain ranges covers 325 topographic maps at 1:50,000 scale with the area of about 130,000km^2, where covers large-area of desert with fixed/fluid dunes, arid salt desert, gobi, badland and yardang landforms. Therefore it is quite difficult for surveying field-works with these kinds of harsh nature environment, meanwhile, poor-texture image areas caused by large areas of desert and gobi result in another difficulties for GCP collection and precise measurement in both image and object space.

Field works such as GCPs surveying and image annotation have been completed in the test areas by over 500 surveyors and 120 vehicles from 7 production units including Shanxi, Heilongjiang and Sichuan Surveying and Mapping Bureau, Chongqing Surveying and Mapping department, and Surveying and Mapping Bureau in Qinghai, Xinjiang and Gansu provinces since 2006. About 700 GCPs and checking points are measured by differential GPS. The measurement accuracy was better than 1m in planimetry and 1.2m in height. The GCPs are well-distributed in the test area, their average ground distance is about 100-150km (in order to ensure reliable GCP at the in-home designed location, two- or three-point layout plan would be utilized at the designed location, and all of them would be recorded in the number of GCPs). According to the coverage of the SPOT-5 HRS satellite images and the distribution of GCPs, also considering the terrain type, block-adjustment of the whole test area will be divided into 4 sub-testfields (Fig. 3): (B1) Testfield in zone of headstream of Three rivers, Tibet Plateau, China: In this area, 13 SPOT-5 HPS stereo image strips which cover about 120,000km^2 are involved. The GCP measurement was completed in 2006, there are large number of GCPs because this area is the first working area of the WChTM project.

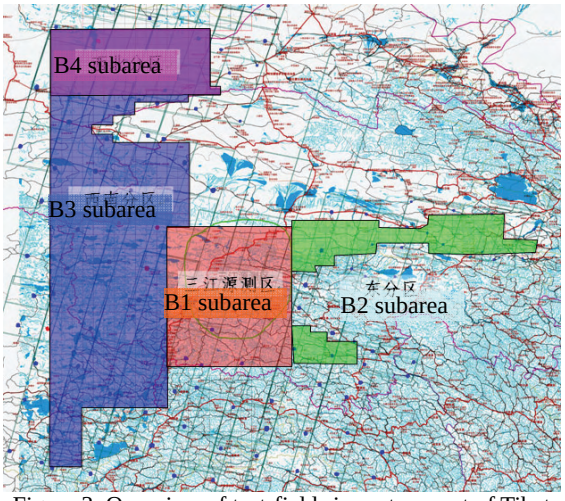


Figure 3. Overview of test-fields in eastern part of Tibet Plateau and eastern Talimu Basin, China

Finally, there are total 157 points were used in block-adjustment procedure except certain number of necessary tie points, in which contain 59 GCPs and 98 checking points. The block-adjustment results are shown in Table 4.

(B2) Test-field of northeast in eastern Tibet Plateau, this area has 26 SPOT-5 HPS stereo image strips which cover about 200,000km^2. There are 209 points were used in block-adjustment for this test-field, in which contain 81 GCPs and 128 checking points. The block-adjustment results are shown in Table 4.

(B3) Test-field of southwest in eastern Tibet Plateau, this area has 26 SPOT-5 HPS stereo image strips with coverage of about 300,000km^2. Finally there are 273 points were used in block-adjustment, in which contain 115 GCPs and 158 checking points. The block-adjustment results are shown in Table 4.

Table 4. Accuracy reports for block-adjustment in test-field of eastern part of Tibet Plateau and eastern part of Talimu Basin, China. The unit is in meters

	σ_0 (pixel)	RMSE of GCPs			RMS of CPs			Max error of GCPs		Max error of CPs	
		X	Y	Z	X	Y	Z	X-Y	Z	X-Y	Z
B1	0.79	4.9	5.2	1.4	8.9	6.9	2.1	13.4	4.3	23.2	4.3
B2	0.68	5.3	4.6	2.1	7.8	6.7	2.5	16.2	5.9	17.4	6.1
B3	0.62	5.2	5.6	1.6	8.4	7.2	2.4	15.8	4.1	18.2	5.0
B4	0.54	4.4	4.5	1.2	5.2	6.3	2.1	11.6	2.2	11.7	3.6

(B4) Test-field of eastern Talimu basin, this area has 5 SPOT-5 HPS stereo image strips with coverage of about 60,000km^2. There are 92 points were used in block-adjustment, in which contain 51 GCPs and 41 checking points. The block-adjustment results are shown in Table 4.

As shown in Table 4, the block-adjustment accuracy of the test areas in eastern Tibet Plateau and eastern Talimu basin are quite good (in sub-pixel level) compared to the resolution of SPOT-5 HRS images, and it can meet the block-adjustment accuracy requirements regulated in mapping standard at 1:50,000 scale with small number of GCPs. Comparing to block-adjustment with the traditional aerial photos, block-adjustment with SPOT-5 HRS stereo image has some different features. For example, the resolution is quite different, the scale of aerial photos used for mapping at 1:50,000 scale normally is 1:35,000-1:50,000, the ground resolution of image usually is in sub-meter or meter level. According to experience of the tests made as above, for block-adjustment with SPOT-5 HRS stereo images, the following problems should be noted specially in GCPs/tie-point measurement, block- adjustment computation and accuracy checking:

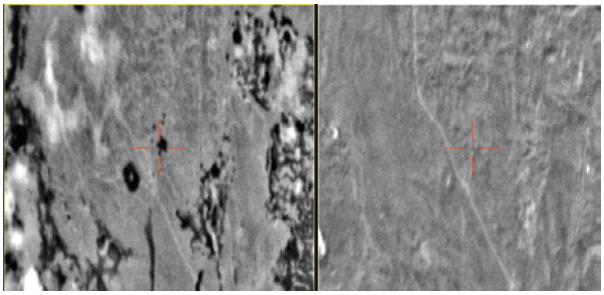


Figure 5. Example of radiometric difference between adjacent SPOT-5 HRS images caused by different acquisition time

(1) For satellite images, the radiometric differences caused by different imaging time between adjacent stereo images must be considered while making the GCP layout plan and measurement. GCPs should be located and can be measured precisely in all (or as more as possible) satellite images. Take example in Fig. 4 for instance, the GCP in left image will be very hardly or even can not be measured/transferred to the adjacent right image if we do not take the temporal difference into account.

(2) For SOPT-5 HRS stereo images, the image resolution of is 5m in along-track and 10m in cross-track direction, this means that measurement accuracy is different in different direction of SPOT-5 HRS images, thus we have to pay much attention when we make the point measurement in cross-track direction.

(3) The requirement of block-adjustment residual errors for GCPs and checking points in existing Chinese mapping standard of aerial image is regulated for tolerance. Sometimes, it might be very difficult to observe image points and carefully modify point location to reach the tolerance requirement. For example, assume the residual error of an image point is 3.2m or 3.5m, but if the tolerance is 3.0m, this means that the modification of this point is in 1/25 or 1/10 pixel level for SPOT-5 HRS image; Therefore it suggests that precise stereo observation must be applied for SPOT-5 HRS images to insure firstly, the residual error of this point in image space is less than 1.5 pixel, meanwhile, it suggests that the residual errors of 5%-10% GCPs or Check points are allowed over requirement of tolerance but they must be less than 1.5 times of the tolerance.

4.2 IRS-P5 Stereo Images and Accuracy Analysis

The test-field is an area around the city of Beijing, China. It consists of a steep mountainous region in the north-western part and flat regions in the middle and southern parts. The city of Beijing is located in the lower part of the study area. The whole area is about 70 × 210 km^2. The site has an average terrain height of 300m and an elevation range of more than 1100 m.

Over the test area, totally 23 scenes of IRS-P5 stereo images were collected. In order to precisely georeference these images, about 66 well-distributed GCPs were collected with differential GPS in 2006. The measurement accuracy was better than 0.5m in planimetry and 1m in height. The GCPs are well-distributed in the test area, their ground intervals are about 30km, and most of them are located at the center of road intersection which can be precisely measurement both in image and object space.

According to works made by Lutes, J. (2006), most orientation errors in IRS-P5 stereo images are either biases or linear in line/sample direction. This suggests that at least 4-6 GCPs are required for orient an IRS-P5 stereo pair (with corresponding RPCs) to achieve good enough results. Other works which made by Jocobsen, et. Al (2008) also show that with model *M_RPC2* (RPCs plus biases correction) can achieve good enough orientation results. However, based on our own test with IRS-P5 images in China, for each scene of stereo image, at least 4 GCPs

are necessary to remove all biases and linear trend errors.

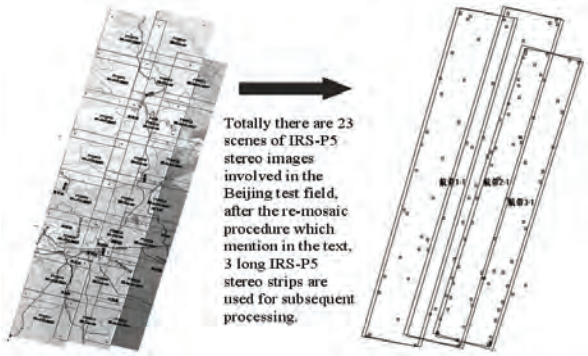


Figure 6. In Beijing test-field, there are 23 scenes of IRS-P5 stereo images. In order to facilitate the subsequent processing, we use the re-mosaic procedure based on metadata to combine these 23 stereo images into 3 long stereo strips.

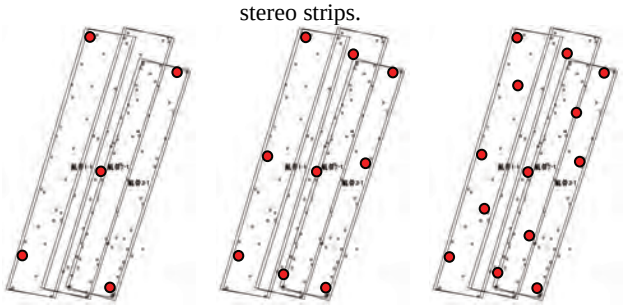


Figure 7. Different distribution of GCPs for different block-adjustment test phase (with 5, 9, 13 GCPs respectively) with IRS-P5 imagery

Normally, we could commercially get so-called IRS-P5 "standard scene" of images, which have 12000 × 12000 pixels and cover 30 × 30km^2. Basically the standard scenes could be sub-images of a long IRS-P5 strip. They are just subdivided from a long IRS-P5 strip for commercial reason and normally they have 5%-15% overlap between adjacent scenes. We develop a procedure to re-mosaic adjacent IRS-P5 standard scenes into a long stereo strip (Fig. 4) in condition that these scenes are sub-images of the original long IRS-P5 strip. After this re-mosaic procedure, 23 IRS-P5 scenes of the Beijing test-field are re-mosaic into only 3 IRS-P5 long stereo strips.

We select different numbers and distribution layout of GCPs for accuracy test, which includes 4 GCP, 9 GCPs, 13 GCPs and all of control points (see Fig. 5), the results are shown in Table 8. Test results show that: (a) with the increase number of GCPs, there is certain degree of improvement for both planimetry and elevation accuracy, but the improvement is not so significant; (b) adjustment accuracy in teat area can meet the requirement of block-adjustment in surveying criterion at 1:50,000 scale even with only 5 GCPs located at corners of the test-field. According to this result, the proposed re-mosaic procedure is quite important for both reduce the number of GCPs and reduce the measurement error accumulation. For single long-strip IRS-P5 images, only 4 GCPs at the image corners are enough to achieve good accuracy under the condition that the length of strip is below 150km; for multiply long-strip IRS-P5 stereo images even less GCPs are required to reach the requirement of topographic map surveying at 1:50,000 scale in China.

Table 8. Accuracy test for block-adjustment in Beijing test-field, with IRS-P5 stereo images. The unit is in meters

Test phase	GCPs+CPs	RMSE-X	RMSE-Y	RMSE-Z
5 GCPs	5 + 61	1.70	1.96	2.49
9 GCPs	9 + 57	1.64	1.93	2.50
13 GCPs	13 + 43	1.61	1.86	2.37
66 GCPs	66 + 0	1.43	1.68	1.98

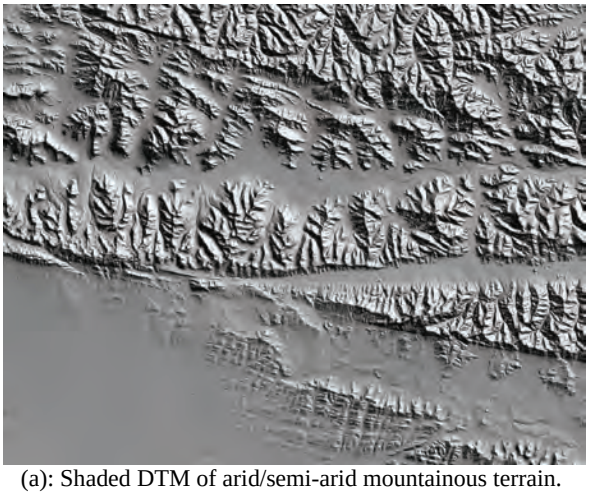
5. PERFORMANCE EVALUATION ON AUTOMATIC DTM GENERATION APPROACH

In order to evaluate the performance of our approach for DTM generation it has been verified extensively with several HRSI datasets, such as IRS-P5 and SPOT-5 HRS/HRG images, over different terrain types, which include hilly and rugged mountainous areas, rural, suburban and urban areas. In the following, we will report in detail about 1 experiment over test area in Zone of headstream of Three rivers, Eastern Tibet Plateau, China. Other processing and evaluation results of IKONOS and SPOT5 HRS/HRG can be found in Poli et al., 2004; Baltsavias et al., 2006 and Poon et al., 2005.

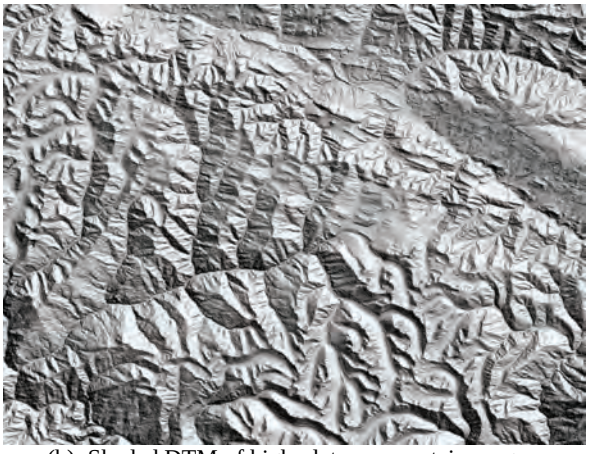
The test area in Zone of headstream of Three rivers, Eastern Tibet Plateau, China covers 250 topographic maps at 1:50,000 scale with the area of about 12,000km^2, where contains large-area of seasonally and perennially frozen soil, mountain/valley glacier and perennial snowfield and large area of unman area. The test-field is the headstream of Yangtze River, Yellow River and Lancangjiang River, and the QingZang railway and national road cross the region from north-east to south-west. The average elevation is 4000m in test-field. Main geological structures are in trend of nearly east-west direction. The various landforms in study area provides better environment for DTM automatic generation.

Over the test area, totally 11 pairs of 10 × 5m SPOT-5 HRS and nearly twenty 5m HRG images were acquired. These images were used to generate DTM over the whole test area. In particular, for DTM accuracy analysis, 2 SPOT-5 HRS satellite image pairs imaged in Nov 2003 and 6 HRG images, which can form the SPOT-5 stereo triplets, have been selected. The images have the fine quality and have no cloud coverage, which provide good data sources for DTM automatic generation and accuracy analysis.

After the block-adjustment, the proposed DTM generation approach was applied to 2 SPOT HRS stereo image strips and 6 HRG images simultaneously. As a result, 25m grid-spacing DTM of the accuracy study area has been generated automatically after about 26 hour. As results, about 90 millions feature points and 5 millions feature lines have been matched which is equivalent to match a 4 × 4 density grid on the original image, and it is good enough for generation of DSM with 25m interval. Since our DSM generation approach not only generates a large number of mass points but also produces line features, which are necessary for the modeling the rugged mountainous terrain. Finally, post-processing of the DSM by using a slope based filtering method was applied and worked well, since the average height of the test area is above 4000m and it is unman area, man-made structures and vegetation don not pose big problems. Parts of DTM results are shown in Fig 6. It can be seen that the resulted DTM reproduced quite well not only the general features of the terrain relief but also small geomorphologic and other features visible in the SPOT-5 images.



(a): Shaded DTM of arid/semi-arid mountainous terrain.



(b): Shaded DTM of high-plateau mountain ranges

Figure 9. The shaded terrain models of 25m over 2 sub-areas in test area. The resulting DTM reproduced quite well not only the general features of the terrain relief but also small geomorphologic and other features visible in the images.

Since the study area is within the WChTM project covered area, there are not good enough reference data, we apply the following three accuracy evaluation methods for DTM accuracy evaluation:

(1) Overlay the automated generated DTM onto the stereo image pairs to apply manual visual checking under stereoscopic display device. Checking results show a good enough match between DTM and stereo images expect some small blunders within shadow areas, which have to apply necessary manual editing.

(2) Accuracy checking by using GCPs and check points measured by differential GPS. Upside of Table 10 give the DTM accuracy evaluation results by using this method. We computed the differences as GCPs minus the interpolated heights from our generated DTM. The accuracy of the DTM is between 1.7 – 6.7 m depending on the terrain relief and land cover.

(3) Accuracy checking by using 1489 manually measured checking points which were acquired on digital photogrammetric workstation. We computed the differences as these points minus the interpolated heights from our generated DTM. The accuracy of the DTM is between 3.6 – 4.5m depending on the terrain relief and landform. Since the DTM generation and point measurement procedure share the same image orientation results, the errors of the DTM would be from the stereo image matching, which is directly relate to the matching accuracy. Detailed accuracies are shown in Table 10 downside.

According to the experiment results, the following conclusions can be made: (a) DTM at 1:50,000 scale can be generated automatically from SPOT-5 HRS/HRG stereo/triplet-stereo images by using the automatic DTM generation approach proposed in this paper. The elevation accuracy can roughly be 50% better than requirements of the Chinese national surveying regulation; (b) simultaneously matching SPOT-5 HRS/HRG triplet stereo strips and matching both feature points and feature-lines are quite necessary for deeply incensed mountainous area and rugged terrain, such as arid and semi-arid broken mountains in northern part and rugged mountain ranges in southern part of the study area; (c) the proposed automatic DTM generation approach can largely increase the working efficiency, however, necessary filtering of the DSM to generate DTMs, careful manually stereoscopic checking and editing are still necessary to remove some small matching blunders, which are in most cases within shadow and gully/rill/ steep-valley well-developed areas.

Table 10: DTM accuracy evaluation results with SPOT-5 HRS/HRG imagery

Reference data	Map sheets	Num. of check points	Height accuracy (RMSE)	Accuracy requirements of 1:50,000 scale $1^{st}/2^{nd}$ level DTM	Type of terrain
GPS points	3	99	6.7 m	14.0 m/19.0 m	Mountainous area
	1	26	3.6 m	8.0 m/11.0 m	Mountainous area
	2	64	1.7 m	5.0 m/ 7.0 m	Hilly area
	3	127	2.8 m	3.0 m/ 4.0 m	Flat/hilly area
Manually measured points	47	937	3.6 m	5.0 m/ 7.0 m	Hilly area
	21	367	4.5 m	8.0 m/11.0 m	Mountainous area
	8	185	4.4 m	14.0 m/19.0 m	Mountainous area

6. CONCLUSIONS

In this paper, we firstly present an approach for block-adjustment based on Rational Function Model (RFM) with sparse GCPs by using satellite Images. To test the proposed approach, it has been applied to SPOT-5 images over test-fields, which covers eastern part of Tibet Plateau, China. The test-fields are with variable terrain geomorphologic type and several tens of GCPs and check points measured by DGPS. The block-adjustment results show that with SPOT-5 HRS images and a small number of GCPs we can achieve 5-9m in planimetric and 2-3m in height direction. In another test we use 23 scenes of IRS-P5 images, the test area covers Beijing area and about 21,000 square kilometers. In this test-field, the block-adjustment result shows that only with 5 GCPs we could achieve 2.0 m in planimetric and 2.5m in height direction. From these experiments, it's shown that with the proposed block-adjustment approach, by using SPOT-5 HRS/HRG and IRS-P5 imagery with several GCPs, satisfactory image orientation results can be completed with a little bit better accuracy than those requirements from Chinese Surveying and Mapping regulations for 1:50000 topographic maps.

Secondly, we have reported about an advanced matching approach for automatic DTM generation from high-resolution satellite images. The proposed approach has been applied to SPOT-5 HRS/HRG images over a test-field in Zone of headstream of Three rivers, Tibet Plateau, China with variable terrain geomorphologic type. The accuracy tests were based on the comparison between as many as 160 accurate GPS check points, more than 1400 manually measured check points and the automatically extracted DTMs. The RMS errors for the whole area are 2-7 m, while for flat/hilly areas the accuracy is about 2-3 m or even better. From the experiment, it's shown that with the proposed automatic DTM generation approach, by using SPOT-5 HRS/HRG and IRS-P5 imagery, satisfactory 1:50000 DTMs can be completed with a better accuracy than those requirements from Chinese Surveying and Mapping regulations.

REFERENCES

Baltsavias, E., Zhang, L., Eisenbeiss, H., 2006. *DSM Generation and Interior Orientation Determination of IKONOS Images Using a Testfield in Switzerland.* Journal of Photogrammetrie, Fernerkundung, Geoinformation, (1), pp. 41-54.

Zhang, L., Gruen, A., 2006. *Multi-image matching for DSM generation from IKONOS imagery.* ISPRS Journal of Photogrammetry and Remote Sensing, 60(2006), pp 195-211

Qian Z., B., Liu J., Y., Xiao G., C., 1990.10, Spaceborne Imagery Photogrammetry. Chinese Liberation Army Press, 270p.

Grodecki, J., Dial, G.. 2003, Block Adjustment of High-Resolution Satellite Images Described by Rational Polynomials. Photogrammetric Engineering and Remote Sensing, Vol. 69, No. 1, pp. 59-68

Jocobsen, et. al., 2008. 8, DEM Generation with CARTOSAT-1 Stereo Imagery. EARSEL Joint Workshop Remote Sensing-New Challenges of High Resolution Satellite Imagery, Bochum.

Lutes, J. First Impression of CARTOSAT-1. 2006, JACIE 2006 Civil Commercial Imagery Evaluation Workshop, Laurel, Maryland, March, 14-16. URL: http//www.geoeye.com/whitepapers_pdfs/2006/ Lutes_CartSat_JACIE2006.pdf

SPOTIMAGING , SPOT Satellite Geometry Handbook, 2002.5, URL: http://www.spotimage.fr//

Tao C. V., Yong H.,A, 2001,67(12), Comprehensive Study of the Rational Function Model for Photogrammetric Processing. Photogrammetric Engineering and Remote Sensing, , pp1347-13

Vosselman, G., 2000, Slope based filtering of laser altimetry data. In ISPRS, Vol.XXXIII B3, Amsterdam, Netherland

FUSION OF ASCENDING AND DESCENDING POLARIMETRIC SAR DATA FOR COLOUR ORTHOPHOTO GENERATION

J. Zhang [a, *], J. Wei [a,b], G. Huang [a], Y. Zhang [a]

[a] Chinese Academy of Surveying and Mapping, Beijing, 100830, P.R.China - zhangjx@casm.ac.cn
[b] Liaoning Technical University, Fuxin, Liaoning Province, 123000, P.R.China

KEY WORDS: SAR, Orthorectification, Polarization, Fusion, Radiometric

ABSTRACT:

Synthetic Aperture Radar (SAR) is an effective earth observation sensor for the regions where frequent cloud and fog cover makes optical image acquisition extremely difficult. The State Bureau of Surveying and Mapping of China is conducting the western China mapping project, which aims to generate 1:50,000 scale topography and other thematic maps for the wild western china region covering around 2 million square kilometers and having not been mapped at this or finer scale before. A part of the western china region characterized by frequent cloud, fog and high mountains will be mapped by SAR. But, due to the inherent side-looking, SAR images suffer strong geometric distortions, such as layover, shadow and foreshortening, especially over mountainous regions. These geometric distortions greatly limited the interpretation of SAR images. To cope with this difficulty, a methodology of generating color SAR orthophotos over mountainous area has been proposed. The methodology consists of several important processes. A fully polarized SAR image is processed to form a color image, which generally has better interpretability than a gray-scale image from single polarized data. The color image is then ortho-rectified with high precision Digital Elevation Model (DEM). At the same time, the radiometric distortion induced by topography, e.g. the variation of effective illumination area within a pixel cell is corrected. Finally, two SAR images with opposite viewing angle, such as the ascending and descending acquisitions, are processed. A mask marking the layover and shadow area is generated based on the DEM. The ascending and descending SAR images are fused by replacing the pixels within the mask with those from the opposite viewing image. Through these processing, a colorful SAR orthophoto is generated, which can be used as a base map for image interpretation. The validity of the proposed methodology has been demonstrated with an experiment using RADARSAT-2 quad-polarization products in the Hengduan mountain area of Western China. A software module has been developed based on the methodology, and has been successfully applied to the Western China Mapping Project.

1. INTRODUCTION

SAR provides an efficient way of earth observation for areas covered by rain, haze/fog and cloud frequently through the year. However, SAR images have severe geometric distortions, such as layover and shadow, which are the consequences of side-looking viewing geometry and underlying topography relief. Precise geometric correction is thus absolutely necessary when spatial information of terrain feature is needed or an integrated analysis of multi-temporal and multi-source information is to be performed (Chen 2004). Since the prevalence of layover and shadow in mountainous areas, scattering signals from these distortion regions hardly contain information about terrain cover types. The State Bureau of Surveying and Mapping (SBSM) of China is conducting the Western China Mapping Project, which aims to generate 1:50,000 scale topography and other thematic maps for the wild western china region covering around 2 million square kilometers and having not been mapped at this or finer scale before. A part of the western china region characterized by frequent cloud, fog and high mountains, for example the Hengduan mountain area, will be mapped by SAR. Obviously, using a single SAR acquisition can not meet the needs of interpreting land cover and other thematic information in these regions. In this paper, a methodology is proposed to generating color SAR orthophotos by fusion two SAR acquisitions with opposite viewing geometry, e.g. the ascending and descending configuration.

* Corresponding author

2. METHODOLOGY OVERVIEW

The proposed methodology consists of several main processing steps: color composition from polarized SAR, ortho-rectification, terrain-induced radiometric correction, fusion of ascending and descending data. The overall flowchart of the methodology is shown in Fig.1. These processing steps are introduced as follows (Figure 1).

2.1 Color Composition of Polarized SAR

A polarimetric SAR measures the microwave reflectivity of a target using quad-polarizations HH, HV, VH, and VV to form a scattering matrix (Lee 1999). Based on the scattering matrix, many methods can be used to form a false color composite. In this work, the Pauli polarimetric decomposition method is used to generate the Red/Green/Blue channels as shown in Equation 1.

$$\begin{aligned} |a|^2 &\rightarrow \text{Blue} \\ |b|^2 &\rightarrow \text{Red} \\ |c|^2 &\rightarrow \text{Green} \end{aligned} \qquad (1)$$

Where $[a \quad b \quad c] = \frac{1}{\sqrt{2}}(S_{HH} + S_{VV} \quad S_{HH} - S_{VV} \quad 2S_{HV})$, and S_{HH}, S_{HV}, S_{VH}, S_{VV} are the elements of scattering matrix.

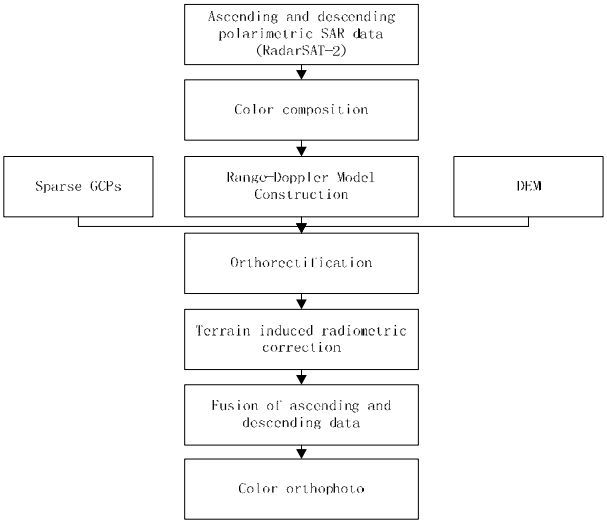


Figure 1. Flowchart of fusion of ascending and descending polarimetric SAR data for color orthophoto generation

2.2 SAR Imagery Ortho-Rectification with Sparse Ground Control Points (GCPs)

The Range-Doppler(R-D) equations (See equation 2) are the physical geometric model for SAR images. For spaceborne SAR, the R-D model can be calculated from the parameters provided in the header file of a SAR image. Based on the model, for every ground resolution cell represented by the 3D geodetic coordinate (X_m, Y_m, H_m), the projected image coordinate (i_c, j_c) can be calculated.

$$\begin{cases} R = |R_s - R_t| \\ f_d = -\dfrac{2}{\lambda}\dfrac{(R_s - R_t)V_s}{|R_s - R_t|} \end{cases} \quad (2)$$

In Equation 2, R_s and V_s are the position vector of the sensor(X_s, Y_s, Z_s)and velocity vector (X_v, Y_v, Z_v) respectively, R_t is the position vector of the ground target (X_t, Y_t, Z_t), R is the distance between the target and the sensor, f_d is the Doppler frequency of the center of the electromagnetic beam and λ is the wavelength of the electromagnetic wave.

But, due to the limited accuracy of the parameters provided in the header file, the calculated image coordinate (i_c, j_c) may offset from the real image coordinate (i_r, j_r), which corresponds to the ground target at (X_m, Y_m, H_m). The relationship between the calculated image coordinate and the real image coordinate can be modeled with a second-order polynomial expressed by equation 3. By means of at least 4 GCPs, these parameters (c_i , d_i) can be estimated using a least-square method. After knowing these parameters, the SAR image can be ortho-rectified to a known DEM (Y. Zhang 2002). Please note, here sparse GCPs are considered, because it is very difficult to obtain many GCPs given the remoteness and inaccessibility of the Western China region.

$$\begin{cases} i_r = c_0 + c_1 \times i_c + c_2 \times i_c^2 \\ j_r = d_0 + d_1 \times j_c + d_2 \times j_c^2 \end{cases} \quad (3)$$

2.3 Terrain Induced Radiometric Correction

It is assumed that the topography is flat in conventional SAR image radiometric calibration. In mountainous areas, the backscattering is distorted by topography relief (Y. Zhang 2003). As for space-borne SAR imagery, the radiometric distortion caused by topography is mainly the variation of effective scattering area. Theoretically speaking, the radiometric distortion caused by the scattering area reach a maximum when the local incident angle is zero. The smaller the incident angle, the bigger the error caused by the topography (Chen 2002). Considering this, we adopt the method proposed by Y. Zhang (2003) to correct the radiometric distortions caused by topography. This method involves calculate a scattering area normalization factor based on DEM and SAR imaging geometry.

The effective scattering area, namely A, is the area a SAR ground resolution cell in actual terrain projected to the normal plane of radar light. It can be calculated in Equation 4 (Adrian 1998).

$$A = \frac{\delta_r}{\sin(\eta_r)}\frac{\delta_a}{\cos(\theta_a)} \quad (4)$$

Where, δ_r is the slant resolution in range direction; δ_a is the azimuth resolution; η_r is the local incident angle in range direction; θ_a is the slope angle in azimuth direction.

As shown in Figure 2, η_r can be calculated according to Equation 5:

$$\eta_r = \gamma - \theta_r \quad (5)$$

where γ is the viewing angle of SAR, and θ_r is the slope angle in range direction.

Since the backscattering coefficient has a square relationship with the amplitude of SAR image, the amplitude value can be multiplied by a scattering area normalization factor σ expressed in equation 6 (Zhang Y 2002) to eliminate the radiometric distortion caused by the variation of effective scattering area of SAR ground cell.

$$\sigma = \sqrt{\sin(\gamma - \theta_r)\cos\theta_a} \quad (6)$$

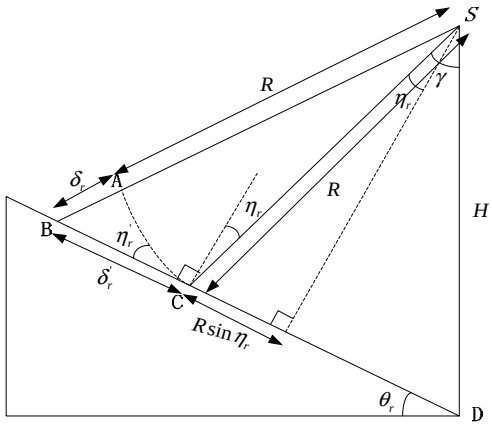


Figure 2. Local incident angle in range direction

2.4 Fusion of Ascending and Descending Images

Based on the terrain slope and aspect angles, a mask image can be generated which marks the areas affected by shadow

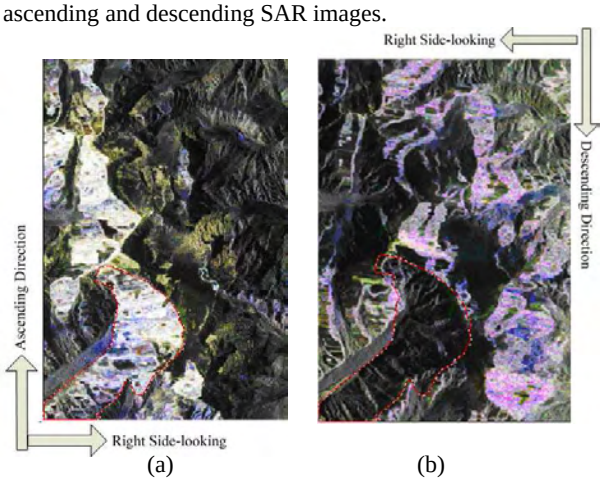
and layover. Since the radar illumination rays of ascending and descending images are opposite to each other, the masks will be complement to each other. Therefore the ascending and descending images can be fused by setting a decision rule to exclude pixels in layover/shadow region. The result image will be full of information and can be used for further interpretation. The details of this processing are introduced in next section.

3. FUSION OF ASCENDING AND DESCENDING SAR IMAGES

3.1 Basic Principle

Because of rough topography, there are severe layover and shadows in ortho-rectified SAR images. Figure 3 shows this phenomenon in ascending and descending SAR images in the Hengduan Mountain areas. Layover and shadow region has very limited useful information, which not only reduces the visual effect, but also results in the difficulty of image interpretation. However, since the lines of sight (LOS) of ascending and descending images are symmetry, the slopes facing the LOS of an ascending image are in the back slopes of corresponding descending image, vice versa. For example, figure 3(a) is a right side-looking SAR image in ascending direction and figure 3(b) is a descending one. The layover area surrounded by the red line in the facing slope in figure 3(a) corresponds to the red area in the back slope in figure 3(b). Therefore, the layover area in figure 3(a) can be substituted by the red area in figure 3(b), and similarly the layover area in figure 3(b) can also be replaced by the corresponding area in the back slope in figure 3(a). By compensating the layover/shadow region of ascending/descending data with the information from counterpart descending/ascending data, a SAR orthophoto with full information can be generated.

Based on the above principle, in this paper a new data fusion method using ascending and descending side-looking SAR images is proposed, including two parts: 1) detection of layover and shadow from SAR images; and 2) data fusion for ascending and descending SAR images.



(a) (b)

Figure 3. (a) Ascending SAR images (b) Descending SAR images

3.2 Detection of layover and shadow

Layover and shadow areas in SAR image need to be detected before fusing ascending and descending data. Based on the SAR geometric model and DEM, a method is developed to detect layover and shadow areas from SAR images. The method includes the following steps:

1) The SAR image with larger incidence angle of the image center is selected as master image, and the other as the slave.

2) Using DEM, the aspect angle in azimuth direction of a pixel in the ortho-rectified master image is calculated, by means of which to determine whether this pixel is located in the facing slope or the back slope.

3) Based on the SAR geometric model, the original row and column coordinates of each pixel in the ortho-rectified master image are calculated. Accordingly, their satellite state vectors can also be obtained by utilizing orbit model.

4) Calculating look angle γ and slope angle θ_r in range direction. If the pixel is located in back slope and satisfies the condition $\theta_r > \gamma$, then it is within shadow area and marked as Flag =1. If the pixel is in the facing slope and meets the condition $\gamma + \theta_r \geq 90^o$, then it is determined to be within the layover area and marked as Flag=0.

5) Generating mask images. The mask images for layover and shadow are generated by conducting the calculation of step 2) ~ 4) for each pixel.

6) Editing the mask images. By means of morphological operations such as erosion, dilation and average filtering, the mask images are edited to eliminate the “islands” and “zigzags”, errors.

The method of detecting for layover and shadow areas mainly involves the following aspects: the judge rules for the facing and back slope, the calculation of radar look angle γ, slope angle in azimuth θ_a and slope angle θ_r in range.

1) Judge rules for the facing and back slope

(1) Heading angle calculation of SAR orbit

Heading angle Ω_s can be calculated with the first and last state vectors which are obtained from the leader file of master image.

(2) Base on DEM, aspect angle β_n of each pixel at ground resolution is calculated.

(3) Calculating the aspect angle in azimuth β_s for each pixel.

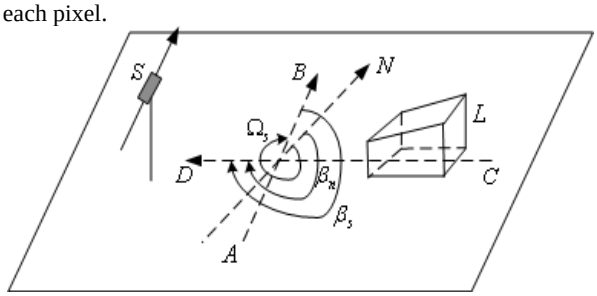


Figure 4. Aspect angle in azimuth direction

The aspect angle in azimuth β_s refers to the angle from flight direction of SAR sensor S to the aspect of ground resolution

pixel L . As shown in Figure 4, $\overrightarrow{AB}$ is parallel to the ascending orbit direction of SAR sensor S , N for the geographic north direction, β_n for the aspect angle of ground resolution pixel L , Ω_s for the heading angle of sensor S . Thus, the aspect angle in azimuth β_s can be expressed as:

a) If aspect angle β_n is greater than heading angle of sensor S , then

$$\beta_s = \beta_n - \Omega_s \quad (7)$$

b) If aspect angle β_n is less than or equal to heading angle of sensor S , then

$$\beta_s = 360 - \Omega_s + \beta_n \quad (8)$$

(4) To determine whether β_s is greater than 180° for each pixel, if YES, the pixel is judged to be in the facing slope, otherwise the back slope.

2) Calculation of radar look angle γ

Based on SAR imaging geometry, radar look angle γ can be expressed as (Jia 2005)

$$\gamma = 90^o - \arccos[\frac{|R_s|^2 + R^2 - |R_T|^2}{2\times|R_s|\times R}] \quad (9)$$

3) Calculation of slope angle θ_a in azimuth and slope angle θ_r in range

θ_a and θ_r are calculated with slope angle, aspect, and heading angle of SAR sensor, which are extracted from DEM.

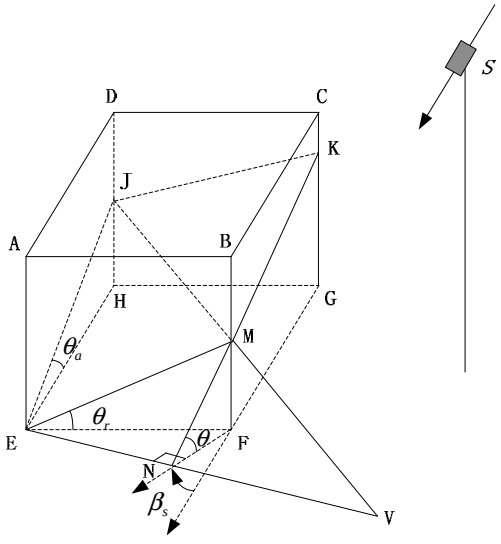


Figure 5. Slope angle θ_a in azimuth and slope angle θ_r in range

As shown in Figure 5, S is the SAR sensor. In the cuboid $ABCD-EFGH$, $\overrightarrow{GF}$ is parallel to the descending azimuth direction, and $\overrightarrow{FN}$ is the aspect of the ground scattering unit $EMKJ$, β_s is the azimuth angle of $\overrightarrow{FN}$ referring to the sensor flight direction $\overrightarrow{GF}$, θ_a and θ_r are the slope angles in azimuth and range of the ground scattering unit $EMKJ$, respectively. According to geometric relationship (Adrian 1998), we have

$$\tan(\theta_r) = \frac{MF}{EF} = \frac{MN\cdot\sin\theta\cdot\sin\beta_s}{MN\cdot\cos\theta} = \frac{\sin\theta\cdot\sin\beta_s}{\cos\theta} \quad (10)$$

Thus,

$$\theta_r = \arctan[\frac{\sin\theta\cdot\sin\beta_s}{\cos\theta}] \quad (11)$$

By the same way,

$$\theta_a = \arctan[\frac{\sin\theta\cdot\cos\beta_s}{\cos\theta}] \quad (12)$$

3.3 Data Fusion of Ascending and Descending SAR Images

The decision-making rule is shown is figure 6, using mask images for layover and shadow, by which the gray value of the pixel in layover and shadow areas can be replaced with corresponding pixel for multi-side-looking SAR data fusion.

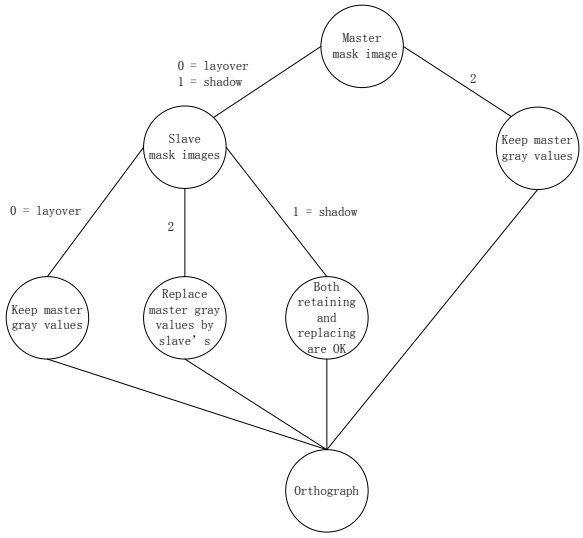


Figure 6. Fusion of ascending and descending SAR images

4. EXPERIMENT ANALYSIS AND APPLICATION

4.1 Test Area and Data

Based on the methodology and algorithms discussed above, a specific module has been developed. An experiment has been conducted using two RadarSat-2 (ascending track and descending track respectively) fully polarized images with 8m resolution in Hengduan Mountain area (see Fig. 7) to test the validity of the proposed methodology and the developed software.

Hengduan mountain area as a whole shows an east-west to north-south direction. It is characterized with high mountains, deep valleys, and complex landform types. The test area has a elevation varying from 1658 meters to 6719 meters above see level and the average 3642 meters. This topography has led to the severe foreshortening, layover and shadows in both the ascending and descending SAR images. The test area has very long time cloudy/foggy weather during a year, and

some meteorological disasters (thunder, snow and hailstone) occurs very often.

The complex topography and harsh natural conditions make traditional field work extremely hard. It is impossible to survey many GCPs in the field .Therefore, some points with clear visual feature at the Radarsat-2 images are seleted from SPOT-5 HRS panchromatic image (5 meters resolution) to be used as GCPs. Totally, 19 points are located with their planimetric coordinates measured from ortho-rectified SPOT 5 image and height from existing 1:100,000 DEM. Out of them, 5 points are used as GCPs to rectify the Radarsat-2 images with the method presented at section 2.2, and the rest 14 points are used as check points. The planimetric positioning accuracy of the rectification in the X direction is about 8m, in the Y direction about 7.5m, and the RMS is around 11m, which fully meets the standards of 1:50000 scale topography map regulated by SBSM.

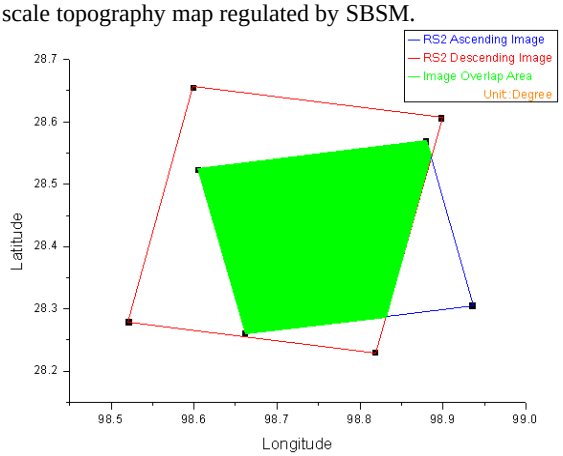


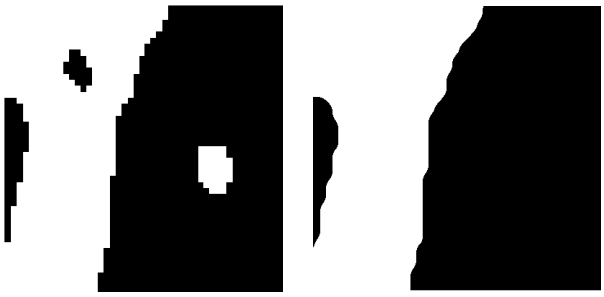
Figure 7. Spatial coverage of the ascending and descending RadarSat-2 images

4.2 Experiment Results and Analysis

4.2.1 Shadow and layover area detection According to the method of shadow and layover detection introduced in section 3.2, the ascending RadarSAT-2 image is chosen as the master image, and the descending one as the slave image. With the existing 1:100,000 DEM of the measured area, the shadow and layover areas of the SAR images are extracted as the mask map (as shown in Figure 8).

Figure 8. Mask map

Due to the limited accuracy of sensor orbit and DEM, for example the DEM we have used is at 1:100,000 scale, which generally has a grid spacing of 50 meters, the extracted mask map will present some errors, such as holes and zigzag. In order to reduce the impact of these errors on the afterward fusion process, the morphological erosion and dilation operations is used to filter the mask image. The effect of the morphological filtering is shown in Fig. 9.



(a) (b)

Figure 9. (a)Original mask with hole and zigzag (b)After morphological filtering

4.2.2 Fusion of ascending and descending SAR images With the mask map of shadow and layover, the ascending image as the master image, the descending one as the subordinate image, the fusion experiment is carried out under integrated decision-making rules. The result is shown in Figure 10.

Figure 10(c) is the resulted SAR color orthophoto after image fusion. Figure 10(a) and Figure 10(b) are the ascending and descending rectified RadarSAT-2 image respectively. These figures show that after image fusion, the shadow and layover areas have been greatly mitigated and eventually a better color SAR orthophoto has been obtained. Moreover, this method is also applicable to the case of fusing moscaiced or subset images from different tracks, to cover a larger or only a portion of SAR frame.

4.3 Application

Based on the proposed methodology of color SAR orthophoto generation, the SAR color orthophotos covering 700,000 square kilometers in the Hengduan Mountain areas have been produced. These orthophotos have been used as the base map of field work to investigate the land cover and thematic information in this mountainous region.

5. DISCUSSION AND CONCLUSION

In Fig10(c), there are still some residual layovers and shadows, which are mainly caused by the following reasons:

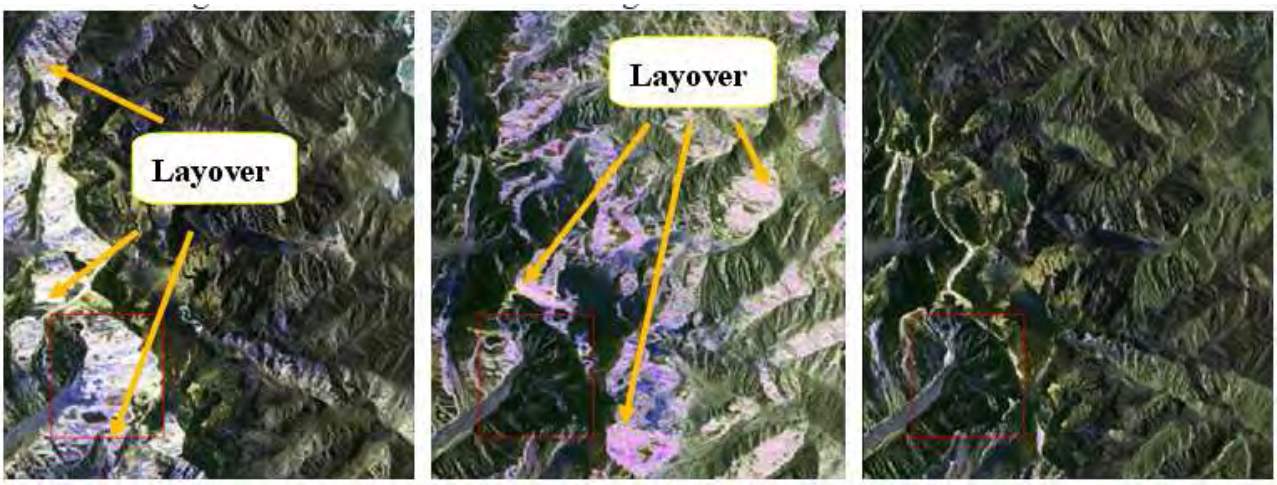
1) The incidence angles of the ascending and descending images are not exactly opposite, e.g. they are 39.02° and 40.53° respectively in the case of Figure 10, leading to that the range and location of layover and shadow areas in the ascending and descending images are not precisely complement to each other. Therefore, we need to carefully choose the ascending and descending images for fusion.

2) The DEM used in this experiment has limited accuracy, which causes the errors in shadow and layover extraction.

But we have to bear in mind that for spaceborne systems the incidence angle of SAR image can not be adjusted as we

want, therefore the first reason may always exist. A solution to this problem is to use more images rather than two with one in ascending track and another in descending track for data fusion. For example, we can use 4 images with each 2 in ascending/descending directions. This can be achieved with spaceborne or airborne SAR systems with the ability of adjustable viewing angles. In this case, it will be not a problem to get a totally layover/shadow free orthophoto using the proposed data fusion scheme.

In response to the complex terrain and intrinsic features of SAR slant imaging in mountainous areas, this paper presents a set of methodology to generate color SAR orthophoto, where the images are acquired in both ascending and descending directions and fused together to exclude pixels within layover and shadow regions. The proposed methodology has been tested and validated and has been put into use in the Western China Mapping Project. However, the methodology can be improved further by using more images to get a totally layover/shaodow free fusion, and balancing the tone differences of the images during different side-looking image fusion.



(a) (b) (c)

Figure 10. (a) Rectified ascending RADARSAT-2 image (b) Rectified descending RADARSAT-2 image (c) Ascending and descending RADARSAT-2 image fusion result

REFERENCES

Luckman, A. J., 1998. Correction of SAR Imagery for Variation in Pixel Scattering Area Caused by Topography. *IEEE Transactions on Geoscience and Remote Sensing*, pp. 344-350.

Erxue Chen, Study on Ortho-rectification Methodology of Space-borne Synthetic Aperture Radar Imagery.Thesis (PHD),2004.47-70. Chinese Academy of Forestry.

Erxue Chen, 2002. Space-borne SAR Image Radiometric Correction Algorithm. *High Technology Letters*, pp. 10-14.

Jong-Sen Lee, Mitchell R. Grunes, Thomas L. Ainsworth , Li-Jen Du and Shane R. Cloude., 1999. Unsupervised Classification Using Polarimetric Decomposition and the Complex Wishart Classifier. *IEEE Transactions on Geoscience and Remote Sensing*, 37(5),pp. 2249-2258.

Y. Zhang, J. Zhang and C. Yang, 2002. Terrain induced SAR radiometric distortions and their corrections. *Science of Surveying and Mapping*, 27(4), pp. 23-26.

Y. Zhang and J. Zhang, 2003. A Quasi-automatic rectification method of SAR image based on Image simulation. *Journal of Remote Sensing*, 7(2), pp. 106-111.

Yonghong Jia, 2005. *Data Fusion Techniques for Multisources Remotely Sensed Imagery*. SinoMaps Press, Beijing, pp. 44-48.

Author Index

Keyword Index