

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

XXXVI

PART
TOME
TEIL

3 / W49B

PIA07

Photogrammetric Image Analysis

Munich, Germany
September 19 – 21, 2007

Part B

Papers accepted on the basis of extended abstracts

Editors

U. Stilla, H. Mayer, F. Rottensteiner, C. Heipke, S. Hinz

Organised by

Institute of Photogrammetry and Cartography,
Technische Universität München

in Cooperation with

ISPRS WG I/2 – SAR and LIDAR Systems
ISPRS WG III/2 – Surface Reconstruction
ISPRS WG III/4 – Automatic Image Interpretation for City-Modelling
ISPRS WG III/5 – Road Extraction and Traffic Monitoring
ISPRS WG IV/3 – Automated Geo-Spatial Data Acquisition and Mapping

This compilation © 2007 by the International Society for Photogrammetry and Remote Sensing. Reproduction of this volume or any parts thereof (excluding short quotations for the use in the preparation of reviews and technical and scientific papers) may be made only after obtaining the specific approval of the publisher. The papers appearing in this volume reflect the authors' opinions. Their inclusion in this publication does not necessarily constitute endorsement by the editors or by the publisher. Authors retain all rights to individual papers.

Cooperating ISPRS Working Groups

- WG I/2 – SAR and LIDAR Systems
- WG III/2 – Surface Reconstruction
- WG III/4 – Automatic Image Interpretation for City-Modelling
- WG III/5 – Road Extraction and Traffic Monitoring
- WG IV/3 – Automated Geo-Spatial Data Acquisition and Mapping

ISPRS Headquarters 2004-2008

c/o ORHAN ALTAN, ISPRS Secretary General
Division of Photogrammetry, Faculty of Civil Engineering
Istanbul Technical University
Ayazaga, 34469 Istanbul, Turkey
Phone: +90 212 285 38 10
FAX: +90 212 285 65 87
Email: oaltan@itu.edu.tr

ISPRS WEB Homepage: <http://www.isprs.org>

Published by

Institute of Photogrammetry and Cartography
Technische Universitaet Muenchen

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

Workshop Committees

Conference Chairs:

Uwe Stilla (Technische Universitaet Muenchen)

Helmut Mayer (Bundeswehr University Munich)

Franz Rottensteiner (University of Melbourne)

Christian Heipke (Leibniz Universität Hannover)

Cooperating Working Groups:

WG I/2 – SAR and LIDAR Systems

WG III/2 – Surface Reconstruction

WG III/4 – Automatic Image Interpretation for City-Modelling

WG III/5 – Road Extraction and Traffic Monitoring

WG IV/3 – Automated Geo-Spatial Data Acquisition and Mapping

Local Organizing Committee:

Stefan Auer (Technische Universitaet Muenchen)

Konrad Eder (Technische Universitaet Muenchen)

Karl Heiko Ellenbeck (University of Bonn)

Christine Elmauer (Technische Universitaet Muenchen)

Stefan Hinz (Technische Universitaet Muenchen)

Ludwig Hoegner (Technische Universitaet Muenchen)

Jens Leitloff (Technische Universitaet Muenchen)

Dominik Lenhart (Technische Universitaet Muenchen)

Manfred Stephani (Technische Universitaet Muenchen)

Program Committee:

Richard Bamler (DLR, Oberpfaffenhofen)
Claus Brenner (Leibniz Universität Hannover)
Joachim Ender (FGAN-FHR, Wachtberg)
Wolfgang Förstner (University of Bonn)
Paolo Gamba (University of Padua)
Marco Gianinetto (Politecnico di Milano)
Norbert Haala (Universität Stuttgart)
Lars Harrie (Lund University)
Christian Heipke (Leibniz Universität Hannover)
Olaf Hellwich (Technical University Berlin)
Stefan Hinz (Technische Universität München)
Jie Jiang (National Geomatics Center of China, Beijing)
Boris Jutzi (FGAN-FOM, Ettlingen)
Peter Krzystek (Munich University of Applied Sciences)
Helmut Mayer (Bundeswehr University Munich)
Chris McGlone (SAIC, Chantilly, VA)
Liqui Meng (Technische Universität München)
Bryan Mercer (Intermap Technologies Corp., Calgary)
Franz Meyer (Alaska Satellite Facility, Fairbanks)
Moritz Neun (University of Zurich)
Kian Pakzad (Leibniz Universität Hannover)
Nicolas Pappas (IGN / MATIS, St. Mandé)
Ingo Petzold (University of Zurich)
Ross Purves (University of Zurich)
Franz Rottensteiner (University of Melbourne)
Daniel Scharstein (Middlebury College)
Jochen Schiewe (University of Osnabrück)
Monika Sester (Leibniz Universität Hannover)
Uwe Soergel (Leibniz Universität Hannover)
Uwe Stilla (Technische Universität München)
Charles Toth (Ohio State University, Columbus)
Markus Ulrich (MVTec Software GmbH, Munich)
Stephan Winter (University of Melbourne)

Preface

Automated extraction of topographic objects from remotely sensed data is an important topic of research in Photogrammetry, Remote Sensing, GIS, and Computer Vision. This joint conference of ISPRS working groups I/2, III/2, III/4, III/5, and IV/3, held at Technische Universitaet Muenchen (TUM), discussed recent developments, the potential of various data sources, and future trends both with respect to sensors and processing techniques in automatic object extraction. The focus of the conference lay on methodological research.

The conference addressed researchers and practitioners from universities, research institutes, industry, government organizations, and private companies. The range of topics covered by the conference is reflected by the cooperating ISPRS working groups:

- SAR and LIDAR Systems (WG I/2)
- Surface Reconstruction (WG III/2)
- Automatic Image Interpretation for City-Modelling (WG III/4)
- Road Extraction and Traffic Monitoring (WG III/5)
- Automated Geo-Spatial Data Acquisition and Mapping (WG IV/3)

Prospective authors were invited to submit a full paper of maximum 6 pages and we received 49 papers for review. The presented papers have undergone a rigorous "double blind" review process of full papers, with a rejection rate of 30%. Each paper was reviewed at least by three members of the program committee. Accepted papers (34) and one invited paper are published as printed proceedings in the IAPRS series as well as on CD labelled as "Part A". Only a subset of these papers could be presented orally due to the single track design of PIA07 and the generous time slots for intensive discussion.

Authors who intended to present application oriented work that was in particular suitable for interactive presentation were invited to submit an extended abstract. A group of the program committee selected 32 out of 55 contributions for presentation. Accepted contributions based on abstract review were invited to submit full papers which are published on CD labelled as "Part B".

In total, we received contributions from authors coming from 26 countries. The proceedings include 66 papers from authors coming from 19 countries. There were 7 oral sessions with altogether 20 papers and two interactive sessions where 46 papers were presented.

Finally, the editors wish to thank all contributing authors and the members of the Program Committee. In addition, we like to express our thanks to the Local Organising Committee, without whom this event could not have taken place. Konrad Eder and Christine Elmauer did a great job in arranging the event. Ludwig Hoegner was very helpful especially with the management of the ConfTool. The final word processing of all incoming manuscripts and the preparation of the CD by Dominik Lenhart is gratefully acknowledged. Jens Leitloff organised the internet connection during the PIA07 event.

Munich, August 2007

Uwe Stilla, Helmut Mayer, Franz Rottensteiner, Christian Heipke, Stefan Hinz

Contents

Automatic Non Parametric Procedures for Terrestrial Laser Point Clouds Processing A. Beinat, F. Crosilla, D. Visintini, F. Sepic <i>University of Udine, Italy</i>	1
Novel Platform for Terrestrial 3D Mapping from Fast Vehicles A. Blug, C. Baulig, M. Dambacher, H. Wölfelschneider, H. Höfler <i>Fraunhofer Institute for Physical Measurement Techniques, Germany</i>	7
An Integrated Approach to Level-of-Detail Building Extraction and Modelling Using Airborne LIDAR an Optical Imagery F. O. Chikomo, J. P. Mills, S. L. Barr <i>FGAN-FOM, Germany;</i> <i>University of Newcastle upon Tyne, United Kingdom</i>	13
A Semi-Automatic High Resolution SAR Data Interpretation Procedure F. Dell'Acqua, P. Gamba, G. Lisini <i>University of Pavia, Italy</i>	19
Object Extraction at Airport Sites Using DTMs/DSMs and Multipectral Image Analysis N. Demir, E. Baltsavias <i>ETH Zürich, Switzerland</i>	25
An Approach for Semi-Automatic Extraction of Features from Aerial Photographs O. Eker, A. C. Kiraci, D. Z. Şeker <i>Harita Genel Komutanlığı, Turkey;</i> <i>İstanbul Teknik Üniversitesi, Turkey</i>	31
Multibaseline Interferometric SAR at Millimeterwaves Test of an Algorithm on Real Data and a Synthetic Scene H. Essen, T. Brehm, S. Boehmsdorff, U. Stilla <i>FGAN-FHR, Germany;</i> <i>Technical Center of the German Armed Forces, Germany;</i> <i>Technische Universitaet Muenchen, Germany</i>	35
A “Sight-Speed” Human-Computer Interaction for Augmented Geospatial Data Acquisition and Processing Systems G. Gienko, E. Levin <i>University of the South Pacific, Fiji Islands;</i> <i>Michigan Technological University, USA</i>	41
Digital Surface Models from UltraCam-X Images A. Gruber, S. Schneider <i>Microsoft Photogrammetry, Austria</i>	47

Close Range Photogrammetry Used for the Monitoring of Harbour Breakwaters	
M. Hennau, M. De Wulf, R. Goossens, D. Van Damme, J. De Rouck, P. Hanssens, L. Van Damme <i>Ghent University, Belgium;</i> <i>VLG-LIN-AWZ-Waterwegen Kust, Belgium;</i> <i>VO-MOW-Maritieme Toegang-Cel Kusthavens, Belgium</i>	53
Automatic Reconstruction of Shape Evolution of ETFE-Foils by Close-Range Photogrammetric Image Analysis	
S. Hinz, M. Stephani, L. Schiemann, F. Rist <i>Technische Universitaet Muenchen, Germany</i>	59
Automated Generation of Building Textures from Infrared Image Sequences	
L. Hoegner, U. Stilla <i>Bundeswehr University Munich, Germany</i>	65
Robust Disparity Maps with Uncertainties for 3D Surface Reconstruction or Ground Motion Inference	
A. Jalobeanu, D. D. Fitzenz <i>CNRS - Université Strasbourg 1, France</i>	71
Extracting Orthogonal Building Objects in Urban Areas from High Resolution Stereo Satellite Image Pairs	
T. Krauß, P. Reinartz, U. Stilla <i>German Aerospace Center (DLR), Germany;</i> <i>Technische Universitaet Muenchen, Germany</i>	77
Automatic Traffic Monitoring with Airborne Wide-Angle Digital Camera System for Estimation of Travel Times	
F. Kurz, B. Charmette, S. Suri, D. Rosenbaum, M. Spangler, A. Leonhardt, M. Bachleitner, R. Stätter, P. Reinartz <i>German Aerospace Center (DLR), Germany;</i> <i>Technische Universitaet Muenchen, Germany;</i> <i>General German Automobile Association (ADAC), Germany</i>	83
Inferring Traffic Activity from Optical Satellite Images	
J. Leitloff, S. Hinz, U. Stilla <i>Technische Universitaet Muenchen, Germany</i>	89
Buildings' Modelling for Digital Orthorectification of Urban Areas Images	
B. D. Marinov <i>University of Architecture, Civil Engineering and Geodesy, Bulgaria</i>	95
Identification of Urban Features Using Object-Oriented Image Analysis	
O. D. Mavrantza, D. P. Argialas <i>Ktimatologio S.A. (Hellenic Cadastral Agency), Greece;</i> <i>National Technical University of Athens, Greece</i>	101

Point Positioning Accuracy of Airborne LIDAR Systems: A Rigorous Analysis N. C. May, C. Toth <i>Ohio State University, USA</i>	107
National and Regional Scale DEMs Created from Airborne InSAR B. Mercer <i>Intermap Technologies Corp., Canada</i>	113
Automatic Class Mean Calculation of Road Surface from IKONOS Images Using Fuzzy Logic and Particle Swarm Optimization A. Mohammadzadeh, M. J.Valadan Zoej, A. Tavakoli <i>K.N.Toosi University of Technology, Iran;</i> <i>Amirkabir University of Technology, Iran</i>	119
Moving Targets Velocity and Direction Estimation by Using a Single Optical VHR Satellite Imagery M. Pesaresi, K. Gutjahr, E. Pagot <i>European Commission, Joint Research Centre, Italy;</i> <i>Joanneum Research Forschungsgesellschaft mbH, Austria</i>	125
Road Junction Extraction from High Resolution Aerial Images M. Ravanbakhsh, C. Heipke, K. Pakzad <i>Leibniz Universität Hannover, Germany</i>	131
Single Tree Detection in Forest Areas with High-Density LIDAR Data J. Reitberger, M. Heurich, P. Krzystek, U. Stilla <i>Munich University of Applied Sciences, Germany;</i> <i>Bavarian Forest National Park, Germany;</i> <i>Technische Universitaet Muenchen, Germany</i>	139
Building Change Detection from Digital Surface Models and Multi-Spectral Images F. Rottensteiner <i>University of Melbourne, Australia</i>	145
Automatic Velocity Estimation of Targets in Dynamic Stereo N. Scherer, M. Kirchhof, R. Schäfer <i>FGAN-FOM, Germany</i>	151
TerraSAR-X Value Added Image Products N. Schmidt, M. Weber, J. Raggam, K. Gutjahr, A. Wimmer <i>Infoterra GmbH, Germany;</i> <i>Joanneum Research Forschungsgesellschaft mbH, Austria</i>	157
Improvement of Interior and Exterior Orientation of the Three Line Camera HRSC with a Simultaneous Adjustment M. Spiegel <i>Technische Universitaet Muenchen, Germany</i>	161

Multi-View Stereo and LIDAR for Outdoor Scene Modelling C. Strecha, W. von Hansen, L. Van Gool, U. Thoennessen <i>KU-Leuven ESAT/PSI, Belgium;</i> <i>FGAN-FOM, Germany;</i> <i>ETH Zürich, Switzerland</i>	167
Using Pavement Markings to Support the QA/QC of LIDAR Data C. Toth, E. Paska, D. Brzezinska <i>Ohio State University, USA</i>	173
Registration of Near Real-Time SAR Images by Image-to-Image Matching B. Wessel, M. Huber, A. Roth <i>German Aerospace Center (DLR), Germany</i>	179
A Multi-Level Image Description Model Based on Digital Topology G. Zhu, X. Liu, Z. Jia, Q. Li <i>Wuhan University, China;</i> <i>China Seismological Bureau, China</i>	185
Keyword Index	191
Author Index	193

AUTOMATIC NON PARAMETRIC PROCEDURES FOR TERRESTRIAL LASER POINT CLOUDS PROCESSING

Alberto Beinat, Fabio Crosilla, Domenico Visintini, Francesco Sepic

Department of Georesources & Territory, University of Udine, via Cotonificio, 114 I-33100 Udine, Italy
 alberto.beinat@uniud.it; fabio.crosilla@uniud.it; domenico.visintini@uniud.it; sepic@dimi.uniud.it

KEY WORDS: Laser scanning, detection, registration, segmentation, classification.

ABSTRACT

The paper deals with the registration and modelling of terrestrial laser point clouds. For both problems a non parametric regression is suitably exploited, whose unknowns are the function values and the partial derivatives of a second order Taylor's expansion estimated for a certain number of surface points. These allow to directly estimate local curvatures, namely *Gaussian*, *mean* and *principal* values. Relating to the registration problem, tie points are automatically detected from point clusters having extreme *Gaussian* curvature values. The centroids of such clusters generate a vertexes configuration: the point to point correspondences are automatically defined by the analysis of the respective adjacency matrices. For these sets of pairs, the pre-alignment roto-translation parameters are computed by a SVD algorithm, while the final alignment is executed by an ICP method. The paper further proposes a method to directly detect the discontinuities (segmentation) and to successively estimate the parameters for each recognized surface (classification). For both goals, the algorithm exploits again the curvature values: the discontinuity contours are characterized by points having *mean* curvature greater than a threshold, while classification is performed by a cluster analysis of points having homogeneous curvature values. Some numerical examples show the proper applicability of the proposed method for coarse and fine registration of different scans, for edge detection, and for surface primitives classification.

1. INTRODUCTION

As well-known registration, segmentation and classification are three main phases of laser data processing once different point clouds are available for a certain object. To automatically carry out these phases in a sequential way, a non parametric analytical technique is proposed in this paper (section 2).

About the registration, a hybrid technique, to automatically execute the alignment of close range point clouds by evidencing their morphological singularities, is presented (section 3). The method is developed by studying the local *Gaussian* curvature values, computed from the partial derivatives of the Taylor's expansion, by running a clustering procedure of points having extreme curvature values, and finally by determining the centroids of each cluster. For every point cloud, the centroids generate a vertexes configuration. Once the centroids set of pairs are identified, the pre-alignment roto-translation is estimated by a SVD algorithm. The refined alignment is completed by a variant of the ICP method.

In order to reconstruct the geometric primitives embedded in the point cloud, the paper further proposes a method to directly detect the discontinuities (segmentation) and to successively estimate the surface primitives of each recognized object (classification). For both aims, the algorithm exploits the local curvature parameters. Slope discontinuities of the surfaces are evidenced by studying the values of the *mean* curvature (section 4). The procedure allows to automatically identify the band of points corresponding to such edges, since characterised by mean curvature values greater than a fixed threshold.

Within each volumetric primitive, a region growing method is accomplished (section 5). Points belonging to the same feature model are characterised, first, by a strict correspondence among the measured height values and the estimated ones, second, for plane surfaces, by a constant value of the first order partial derivatives, and third, for curvilinear surfaces, by a constant value of the *Gaussian* and *mean* curvatures. Finally a *forward search* procedure allows a robust and refined classification of the remaining points belonging to the studied primitives.

Some numerical examples show the proper applicability of the proposed method with simulated and real noisy data (section 6).

2. APPLICATION OF A NON PARAMETRIC REGRESSION MODEL

As already mentioned, the fundamental steps of the cloud registration and the shape reconstruction mainly exploit the same analytical model based on a nonparametric regression. The main advantage of this approach consists in its full generality, since it does not require any a priori knowledge of the point geometry or the analytical function of interpolation.

2.1. Estimation of local surface parameters

Let us consider the following polynomial model of second order terms (Cazals and Pouget, 2003):

$$Z_j = a_0 + a_1u + a_2v + a_3u^2 + a_4uv + a_5v^2 + \varepsilon_j \quad (1)$$

where the coefficients and the parameters are locally related to a measured value Z_j by a Taylor's expansion of the function $Z = \mu + \varepsilon$ in a neighbourhood point i of j , as:

$$a_0 = Z_{0i}; \quad a_1 = \left(\frac{\partial Z}{\partial X} \right)_{X_i}; \quad a_2 = \left(\frac{\partial Z}{\partial Y} \right)_{Y_i}; \quad a_3 = \frac{1}{2} \left(\frac{\partial^2 Z}{\partial X^2} \right)_{X_i};$$

$$a_4 = \left(\frac{\partial^2 Z}{\partial X \partial Y} \right)_{X_i, Y_i}; \quad a_5 = \frac{1}{2} \left(\frac{\partial^2 Z}{\partial Y^2} \right)_{Y_i}; \quad u = (X_j - X_i); \quad v = (Y_j - Y_i)$$

with X_i, Y_i and X_j, Y_j plane coordinates of points i and j .

The parameters a_i ($i \neq 0$) are the first and second order partial derivatives along X, Y directions in the i -th point of the best interpolating local surface, collected in the vector:

$$\beta = [a_0 \ a_1 \ a_2 \ a_3 \ a_4 \ a_5]^T$$

where a_0 is the estimated function value at point i .

The weighted least squares estimate of the unknown vector β from a limited number of p neighbour points results as:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{Q} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Q} \mathbf{z} \quad (2)$$

where (for $j = 1, \dots, p$):

- \mathbf{X} is the coefficient matrix, with p rows as:

$$\mathbf{X}_j = \begin{bmatrix} 1 & u & v & u^2 & uv & v^2 \end{bmatrix}$$

- \mathbf{Q} is a diagonal weight matrix defined by a symmetric kernel function centred at the i -th point, with diagonal elements as:

$$w_{ij} = \begin{cases} [1 - (d_{ij}/b)^3]^3 & \text{for } d_{ij}/b < 1 \\ w_{ij} = 0 & \text{for } d_{ij}/b \geq 1 \end{cases}$$

where d_{ij} is the distance between the points i, j and b is the half radius of the window encompassing the p closest points to i . The value of b , rather than the kernel function, is critical for the quality in estimating β . In fact, the greater is the value of b , the smoother the regression function results, while the smaller is the value of b , the larger is the variance of the estimated value. As last remark, to apply the second order Taylor expansion (1), the coordinate Z must be univocally defined by X, Y coordinates. This is always true for aerial laser data, where the laser strips are all aligned and geo-referenced e.g. in E, N, H coordinates, while for terrestrial data the scans geometry is more complex. Each point cloud has a different local X, Y, Z reference system, the scans can be panoramic in azimuthal plane and, for some laser devices, quasi-panoramic in the zenithal plane, i.e. with points in the entire direction sphere. To avoid ambiguous cases, it is thus necessary to share the cloud in more sub-clouds. Moreover, it could be necessary a permutation among X, Y, Z coordinates in order to assume, as Z -axis for (1), the direction whose Z values results better expressed as function of the X, Y ones. In other words, points displaced onto quasi "vertical" surfaces are not well modelled by expansion (1).

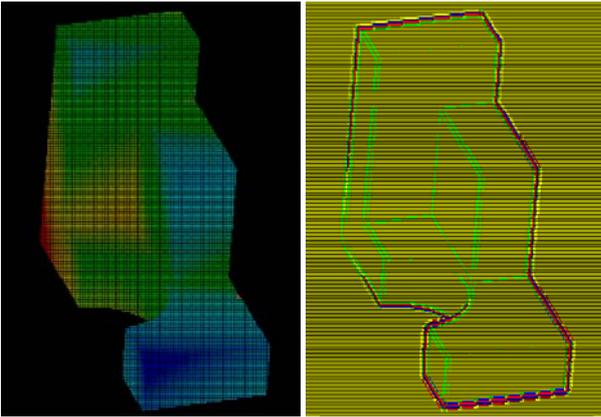


Figure 1: Simulated laser points of the *curvblock-1* model coloured by Z_i values (at left) and by $Z_i - Z_{0_i}$ (at right).

Figure 1 reports the simulated scan *curvblock-1* as example throughout the paper: it comes from the OSU Range Image database (Ohio State University) and has been also experimented by Alshawabkeh, Haala and Fritsch (2006). In Figure 1 at left, the points are coloured by the original Z_i values, since no coordinate permutation is required. At right, the same points are coloured by $Z_i - Z_{0_i}$ absolute values, explainable as the smoothing effect of the regressive

interpolation: the external edges of the object clearly appear, while the internal ones are not significantly recognised.

2.2 Computation of local curvatures values

For the local analysis of a surface S obtained from a point cloud, a great support is provided by geometric quantities coming from differential geometry, in particular by local *Gaussian*, *mean* and *principal* curvatures values. These ones can be obtained from the so-called "*Weingarten map*" matrix \mathbf{A} of the surface S (e.g. Do Carmo, 1976), that is given by:

$$\mathbf{A} = - \begin{bmatrix} e & f \\ f & g \end{bmatrix} \begin{bmatrix} E & F \\ F & G \end{bmatrix}^{-1} \quad (3)$$

where E, F , and G are the coefficients of the so-called "*first fundamental form*", obtained from the parameters estimated by (2) as follows:

$$E = 1 + a_1^2; \quad F = a_1 a_2; \quad G = 1 + a_2^2$$

e, f , and g are the "*second fundamental form*" coefficients, as:

$$e = 2a_3 / \sqrt{a_1^2 + 1 + a_2^2}; \quad f = a_4 / \sqrt{a_1^2 + 1 + a_2^2}; \\ g = 2a_5 / \sqrt{a_1^2 + 1 + a_2^2}$$

The *Gaussian* curvature K corresponds to the determinant of \mathbf{A} :

$$K = \frac{eg - f^2}{EG - F^2} \quad (4)$$

The *mean* curvature H can be instead obtained from:

$$H = \frac{eG - 2fF + gE}{2(EG - F^2)} \quad (5)$$

The *principal* curvatures k_{\max} and k_{\min} , corresponding to the eigenvalues of \mathbf{A} , are given instead from the solution of the system $k^2 - 2Hk + K = 0$, i.e. from $k_{\min, \max} = H \pm \sqrt{H^2 - K}$.

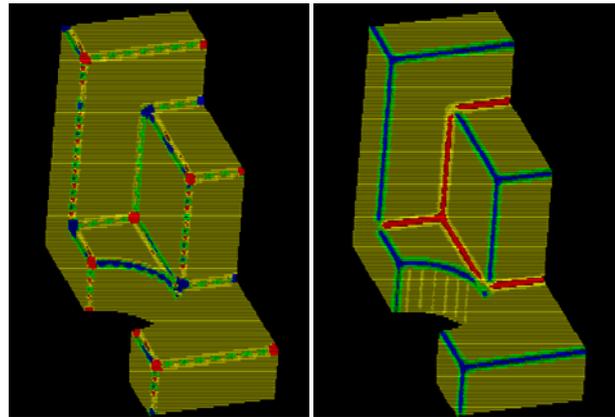


Figure 2: Points coloured by K (at left) and by H values (at right).

Further usable relationships for the curvature values are: $K = k_{\min} k_{\max}$ and $H = (k_{\min} + k_{\max})/2$.

It is interesting to observe in Figure 2 how, thanks to the different geometrical meaning, the *Gaussian* curvature K (at left) can be mainly exploited for the point clouds registration, while the *mean* curvature H (at right) for the edge detection.

Summarizing, for each i -th laser point, four local curvature

values K , H , k_{\max} and k_{\min} can be automatically obtained as functions of the vector β terms. Furthermore, such curvatures are invariant to the reference frame rotations, providing a very important property in analyzing the surface shape.

3. POINT CLOUDS REGISTRATION

3.1 Vertexes detection by means of the K curvature values

Once the *Gaussian* curvature K values are determined by (4) for all the sampled points, it is worthwhile to consider those having extreme absolute values, represented in Figure 2 at left with blue and red dots: they are the “vertexes” of the scanned object. Such points, automatically detected, can be exploited as “tie points” for registration purposes. More properly, since clusters of points with extreme K values will be found, the mean of their 3D coordinates (the centroid) is considered. The set of thus determined centroids constitutes the first point configuration to submit to a correspondence search. This process is repeated for all the clouds that have to be registered, defining in this way a series of centroid configurations.

3.2 Automatic feature matching and labelling

The centroids identified at step 3.1 form a set of possible candidates to be homologous points of adjacent clouds. The next step consists in the recognition of topological relationships existing among the clusters (labelling problem).

Let us consider two partially overlapped point clouds, from which two sets \mathbf{p} and \mathbf{q} , respectively constituted by m and n tie points, have been individuated, as represented in Figure 3.

The problem consists in defining the intersection $\mathbf{p} \cap \mathbf{q}$, and in automatically finding out, within the intersection, the probable correspondences between the tie points of the sets \mathbf{p} and \mathbf{q} .

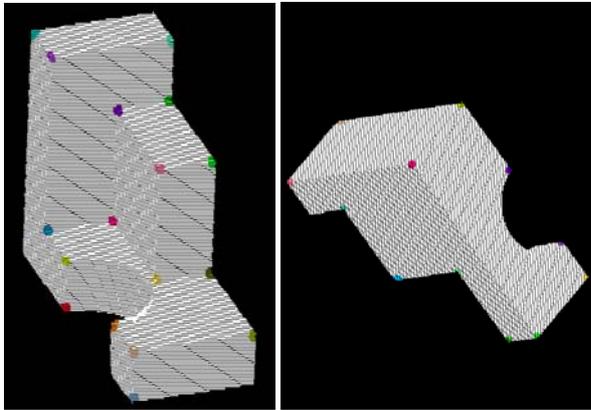


Figure 3: Vertexes detected for *curvblock-1* and *curvblock-2*.

First of all, we assume that no scale variation exists between the coordinate systems of \mathbf{p} and \mathbf{q} . This simplified hypothesis is correct according to the purpose of this operation. The implemented method runs in the following way:

1. Let us consider $\mathbf{p} = \{p_1 \dots p_m\}$ the arbitrary m points configuration. The $m \times m$ symmetric adjacency matrix \mathbf{D}^p of this configuration contains among its terms the Euclidean distance $d_{i,j}^p = \|p_i - p_j\|$ between points p_i and p_j .
2. In the same way, for the arbitrary n points configuration $\mathbf{q} = \{q_1 \dots q_n\}$, the $n \times n$ adjacency matrix \mathbf{D}^q is defined.
3. The row of maximal asymmetry $\mathbf{d}_{i_{\max}}^p$ is searched in \mathbf{D}^p

(or in \mathbf{D}^q), whose distinct elements, ordered in terms of magnitude, present the “maximal minimal difference”, i.e.:

$$\mathbf{d}_{i_{\max}}^p := \left\{ d_{i,j}^p \geq d_{i,j+1}^p; \max_{i=1 \dots m} \left[\min_{j=1 \dots m-1} (d_{i,j}^p - d_{i,j+1}^p) \right] \in \mathbf{d}_{i_{\max}}^p \right\}$$

This search minimizes the possibility of ambiguous geometrical configurations.

4. In the row $\mathbf{d}_{i_{\max}}^p$ of maximal asymmetry, the greatest element $d_{i,j_{\max}}^p = \max \left[\mathbf{d}_{i_{\max}}^p \right]$ is identified. Next step consists in searching $d_{k,l}^p := \left\{ d_{i,j_{\max}}^p - d_{k,l}^p \leq \varepsilon; \forall k, \forall l < k \right\}$ in \mathbf{D}^q , where ε is a prefixed tolerance. The satisfying values k, l are stored into a pointer array, i.e. the list of the possible pairs (q_k, q_l) corresponding to $(p_i, p_j)_{\max}$. If this set is empty, the search is repeated considering the next component to $\mathbf{d}_{i,j_{\max}}^p$ in terms of magnitude.
5. The various pairs of possible correspondences (q_k, q_l) are orderly considered. For each row of \mathbf{D}^q , where one of the correspondence pairs is present, the equivalence of the remaining elements of the same row with respect to the possible elements of $\mathbf{d}_{i,j_{\max}}^p$, within a fixed tolerance, is verified. This allows to generate a binary table, of size $m \times n$, where the elements express the possible correspondence among the points of \mathbf{p} and of \mathbf{q} , according to the initial choice for i_{\max} and k , respectively.
6. If this table has at least two non null elements, a cross-validation of all the possible correspondences is carried out. This is performed verifying the equivalence among all the remaining distances defined by point pairs of \mathbf{p} and analogous point pairs of \mathbf{q} , inserted into the table.
7. This process is repeated for each pair (q_k, q_l) identified at sub-step 4, adopting the pair generating the largest number of valid correspondences between \mathbf{p} and \mathbf{q} .

A set of implemented tests, makes it possible to solve ambiguous situations. In the next version, to evaluate the correspondence point degree, some attributes associable to the points (e.g. curvatures), will be employed.

3.3 SVD pre-registration

Thanks to the pairs of tie points, identified and related by steps 3.1 and 3.2, two matrices of corresponding point coordinates are obtained. Translations and rotations (without the scale factor), to transform the coordinates of a point cloud onto another one, are determined by applying the SVD method.

3.4 Registration refinement by ICP

Once the pre-registration is obtained, the process is completed by ICP. In our case, a basic version of the method proposed by Besl and Mc Kay (1992) has been implemented, updated by some variants proposed in the literature.

The procedure is here described in the essential way, referring to Beinat, Crosilla and Sepic (2006) for the details.

In performing the registration of three *curvblock* scans, after the detection 3.1 of the tie points reported in Figure 3, the point matching 3.2, the pre-registration 3.3, and the refined registration 3.4 have provided a global model with a very good congruence. It is represented in Figure 4, where *curvblock-1* is coloured in blue, *curvblock-2* in green, and *curvblock-4* in red.

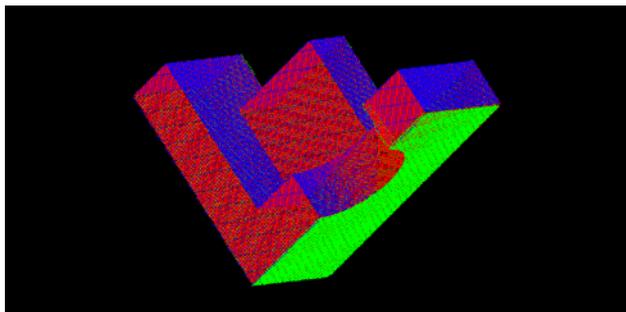


Figure 4: Resulting *curvblock* from a three clouds registration.

4. POINT CLOUD EDGE DETECTION

Once the laser points have been merged into a unique cloud, the incoming problem is the extrapolation of the geometric shape of the surveyed objects with the maximum level of automation. This topic represents the main challenge dealing with the laser scanning technique, namely the maintenance in the data processing of the extraordinary automation level of the data acquisition. In this sense, the different expressions used in literature as edge detection, object recognition, shape classification, and surface reconstruction can be considered as complementary approaches of the same more general problem, defined as “automatic interpretation” of the laser data. This is not a trivial topic and the difficulty grows if we consider noisy mid-long range laser data rather than precise close range ones. In addition, scenes as typical in architectural surveying are more complex than in the industrial environments, since lots of single objects or irregular surfaces occur.

4.1 Edge detection problems: kinds of discontinuity

From the mathematical point of view, the discontinuities of a $Z = f(X, Y)$ surface can be as follows:

- D0: Step discontinuity: the Z values of a significant number of points displaced along a certain X, Y direction present a jump (i.e. the $C0$ continuity is not fulfilled);
- D1: Slope discontinuity: the inclination values of a $C0$ surface change locally ($C1$ continuity is not satisfied);
- D2: Curvature discontinuity: one of the principal curvatures of a $C1$ surface changes locally ($C2$ continuity is not granted).

The difficulty in detecting such discontinuities increases with the rank, anyway the term “edge detection” is mainly thought as the D1 slope discontinuities search. Furthermore, dealing with objects surveyed from a lot of scan positions, the happening of D0 discontinuities normally disappear.

4.2 Analytical and geometrical methods of edge detection

From the methodological point of view, the edge detection problem can be carried out by (at least) three methods, classified as “analytical direct”, “analytical indirect” and “geometrical by decimation”.

The algorithms involving surface interpolations by any analytical function belong to the first class. These have the common property to provide one or more local numerical values directly revealing singularities in the laser cloud. Interesting models have been proposed by the research groups of the Technical University of Wien (Briese, 2006) and the University of Stuttgart (Alshwabkeh, Haala and Fritsch, 2006). The methods belonging to the second family deal instead, first of all, with the suitable estimation of continuous surfaces better interpolating the laser cloud. Only in a second step, the D1 and

D2 edges are detected by considering the space intersection of such surfaces or simply analyzing to which surface each point has been assigned. An analytical procedure proposed by the authors for the classification and the segmentation of laser data (Crosilla, Visintini and Sepic, 2005, 2007) belongs to this class. As general consideration about direct or indirect methods, their most critical characteristics are the requirement of high quality laser data, falling down the efficiency in presence of noisy data, and the modeling complication when a large numbers of parameters have to be estimated.

With a geometric approach, implemented in many commercial software of solid modelling, a TIN mesh is engaged. In this way, the numerical processes applied, as the smoothing and the decimation, regards the optimization of the mesh and does not involve the coordinate points; so this method can be defined as “geometrical”. For this reason, the D1 edge detection is not thus a straight goal of this approach, anyway the edges are strongly correlated with the result of a process of vertex decimation: in fact, they well correspond to the so-called “feature edges” of the triangles remaining after the decimation. Interesting methods are reported in computer vision literature (e.g. Garland, 1997).

As reported before, the analytical indirect method proposed, allows detecting the surface primitives from an undistinguished laser cloud. It proceeds in the four following steps:

1. Estimation of a local surface by a non parametric Taylor’s expansion (as seen in subsection 2.1);
2. Computation of *Gaussian K*, *mean H* and *principal* k_{\max} and k_{\min} local curvatures (as reported in subsection 2.2);
3. Raw segmentation of the cloud in homogeneous clusters by a region-growing method considering also curvature values;
4. Refined segmentation of the raw clusters by a robust parametric regression for each cluster, so estimating the parameters of the various interpolating surfaces.

Focusing the attention to the detection of the three kinds of discontinuity, the following strategy is proposed:

- D0 edges: by checking if the $Z_i - Z_{0_i}$ absolute value is greater than a fixed threshold, as seen in Figure 1 at right.
- D1 edges: by evaluating if the H absolute value is greater than a fixed threshold; the threshold value, as for D0, is fixed considering the noise and the density of the data.
- D2 edges: by estimating the surfaces and by considering the sign and the values variation of K and H .

Nevertheless, the points detected in the previous way must be geometrically significant, that is a certain number of points displaced within a small buffer volume lengthened in one direction should be found. Furthermore, to transform such points in a vector 3D polyline, a suitable chaining or a space interpolation has to be applied.

A rearrangement of the four steps method is now proposed. To detect D1 edges, one more analysis is performed after the second step (see subsection 4.3). In this way, the procedure becomes a direct method of edge detection. Steps 3. and 4. are instead carried out to detect D2 edges, since these curvature discontinuities are found in indirect mode by applying a robust parametric model (see section 5): thanks to this approach, the reliability of the achieved detection should be satisfactory.

4.3 D1 detection by means of the H curvature values

The attention is now focused onto the estimated values of the local *mean* curvature H . The analysis of H values, proposed by the Stuttgart School (Alshwabkeh, Haala and Fritsch, 2006) exploits the property that such index is closely related to the first variation (slope) of a surface area that locally well reveals

possible D1 edges. Since H is the average of k_{max} and k_{min} , it is numerically slightly less sensitive to the noise with respect to K curvature, which is instead the product of k_{max} and k_{min} . Extreme absolute values of H are therefore searched: this point buffer volume so reveals the D1 edges we are looking for.

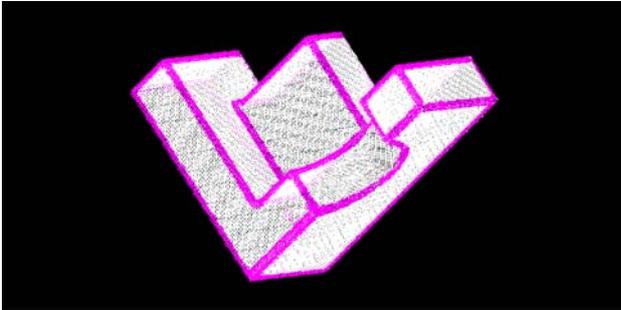


Figure 5: Extreme values H corresponding to possible edges.

This criterion has been implemented in a routine that paints by pink colour the points with H absolute value greater than a threshold, so evidencing the D1 edge zones, as in Figure 5.

5. SURFACE SEGMENTATION AND CLASSIFICATION

As mentioned before, D2 edges are the most difficult elements to automatically detected, especially with noisy data. To this aim and when the main interest of the surveyor is anyway about the overall surface rather than its local discontinuities, efficient methods for its segmentation and classification assume the main importance in the whole flow-chart of data processing.

5.1. Raw surface segmentation by a region-growing method

Analyzing the sign and the values of K and H , a preliminary clustering of the whole cloud is made possible. Each surface can be classified into one of the following basic types (see Table 6): hyperbolic ($K < 0$), parabolic ($K = 0$ but $H \neq 0$), planar ($K = H = 0$), and elliptic ($K > 0$).

	$K < 0$: hyperbolic	$K = 0$: parabolic/planar	$K > 0$: elliptic
$H < 0$			
$H = 0$			not possible
$H > 0$			

Table 6. Classification of surfaces according to the values of Gaussian K and mean H curvatures (from Haala et al., 2004).

This allows to classify the various volumetric primitives and to define a priori the polynomial degree of the parametric model to apply for the refined segmentation (step 5.2).

Hence, to classify and segment the dataset, a region growing method is applied, starting from a random point not yet belonging to any subset. The surrounding points having a distance less than the bandwidth b are analysed, by evaluating the values of the estimated height Z_{0_i} and the values of K and H . If the neighbour points present difference values within a threshold, fixed according to the noise level, then they are labelled as belonging to the same class and putted into a list. The same algorithm is repeated for each list element, till this is fully completed. Afterwards, the procedure restarts again from a new random point, ending when every point has been analysed. Summarising, a first raw segmentation of the whole dataset is carried out in this way: hence, each cluster represents an initial subset to submit to the next refining segmentation step 5.2.

5.2. Refined surface segmentation by a parametric model

Previous authors papers report in detail this step (Crosilla, Visintini and Sepic, 2005, 2007): it is based on the application of a Simultaneous AutoRegressive (SAR) model to describe the trend surface of each point cluster, and on an iterative *Forward Search* (FS) algorithm (Cerioli and Riani, 2003) to find out outliers. Starting from a cluster detected as in step 5.1, the FS approach allows a robust estimation of the SAR unknown parameters. At each iteration, one or more points are joined according to their agreement with the surface model. If some statistical diagnostics reveal an incoming outlier, the growing process is interrupted: the surface is so bounded, hence a refined segmentation is achieved.

Figure 7 shows for *curvblock* point cloud the correct result of the classification, that is one cylinder face ($K=0, H>0$), twelve planes ($K=0, H=0$), and the refined segmentation of the model.

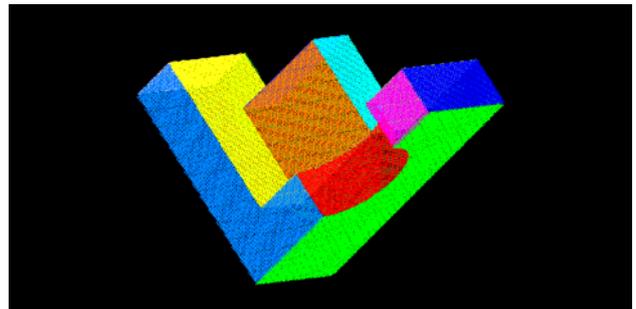


Figure 7: Classification and refined segmentation of the different surfaces: cylinder in red, plane faces in other colours.

6. NUMERICAL EXPERIMENTS

The numerical testing of the proposed procedures has been carried out with satisfactory results for the *curvblock* model scans and for other synthetic objects of the OSU Range Image database. Only some brief comments about the two models depicted in Figure 8 are reported in this paper.

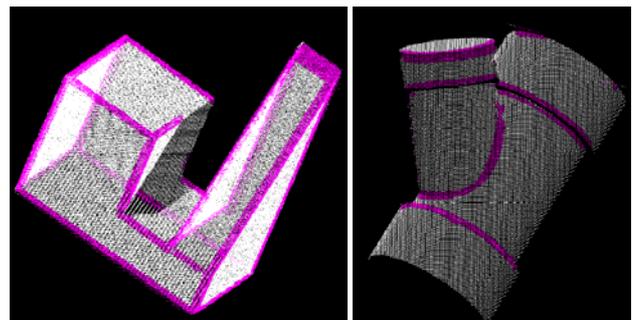


Figure 8: D1 Edges detected for *block2* and for *bigweye* model

The *block2* model (at left) has been joined by registering five partial clouds, exploiting its various vertexes detected by K values analysis. The global point cloud results correctly registered, anyway a central part without points remains for the incompleteness of the data, so yielding a D0 discontinuity. The detection of D1 edges by evaluating H values is completely fulfilled and, since all the surfaces are planes, steps 5.1 and 5.2 have not been carried out.

Model *bigweye* (at right) is instead constituted by curved surfaces: the D1 detection has been correctly accomplished as well as the classification 5.1 as cylinders with values $H < 0$.

Last but not least, to test the method in noisy conditions, some experiments have been performed onto real data acquired with a

Riegl Z360i laser system onto the façade of the baroque Church of Saint Ignatius in Gorizia (Italy). Figure 9 evidences, within more than 500.000 laser points, those having extreme K values, well corresponding to vertexes of the architectonic elements of the façade, and exploitable for registration purposes.

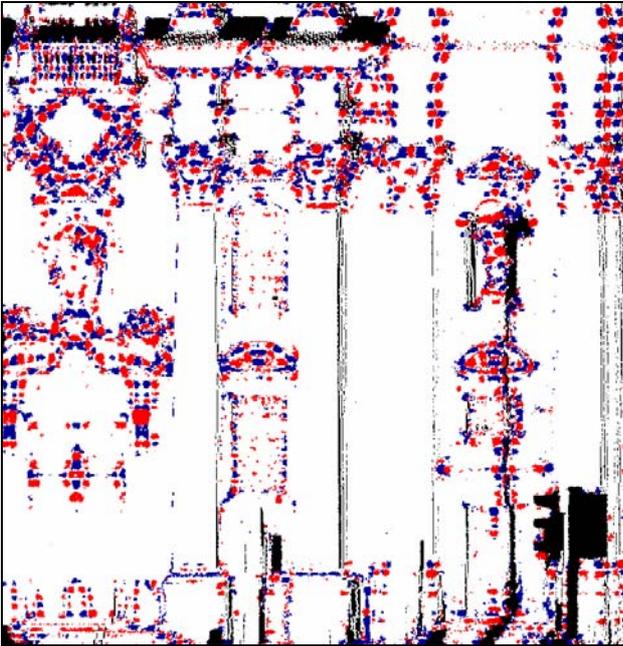


Figure 9: Real points with extreme K values (blue < 0 , red > 0).

Figure 10 shows instead the points with extreme H values, well congruent with the vertical and horizontal edges of the façade.

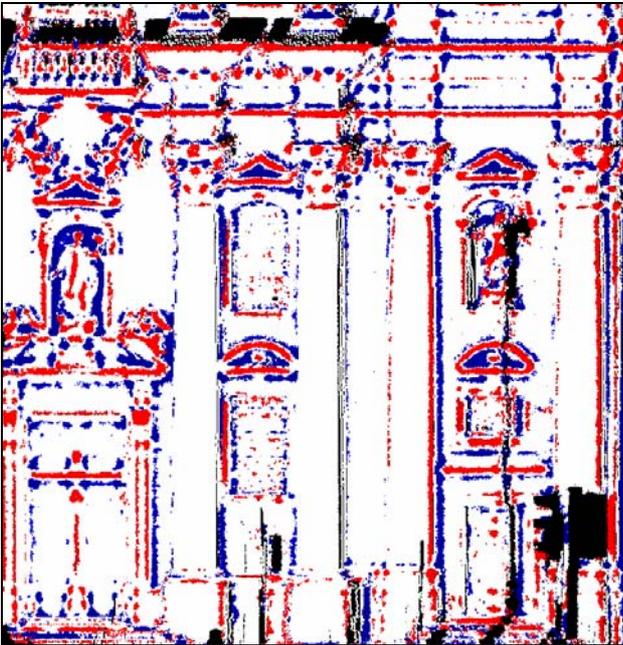


Figure 10: Real points with extreme H values (blue < 0 , red > 0).

The results of this D1 edge detection appear quite satisfactory, particularly in comparison with those geometrically achieved by commercial software implementing TIN decimation tools.

7. CONCLUSIONS

The paper reports a sequential non parametric procedure, mainly based on local estimation of second order Taylor's

expansion partial derivatives, to automatically perform registration, segmentation and classification of laser clouds.

The estimated curvature values allow the automatic detection of singularities in the clouds: registration tie points are evidenced by analyzing the local *Gaussian* curvatures, while segmentation edge points result by evaluating the *mean* curvatures.

The obtained results, for synthetic and real noisy laser data, emphasize the capability of the method proposed for the processing of terrestrial laser surveys.

REFERENCES

- Alshwabkeh, Y., Haala, N., Fritsch, D., 2006. 2D-3D feature extraction and registration of real world scenes, in: *IAPRS&SIS*, XXXVI, 5, Dresden, 32-37.
- Beinat, A., Crosilla, F., Sepic, F., 2006. Automatic morphological pre-alignment and global hybrid registration of close range images, in: *IAPRS&SIS*, XXXVI, 5, Dresden (on CD).
- Besl, P. J., McKay, N. D., 1992. A method for registration of 3-D shapes, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14, 239-256.
- Briese, C., 2006. Structure line modelling based on terrestrial laser scanner data, in: *IAPRS&SIS*, XXXVI, 5, Dresden (on CD).
- Cazals, F., Pouget, M., 2003. Estimating differential quantities using polynomial fitting of osculating jets. in: *Proceedings of the 1st Symposium on Geometry Processing*, 177-187.
- Cerioli, A., Riani, M., 2003. Robust methods for the analysis of spatially autocorrelated data. *Statistical Methods & Applications*, 11, 334-358.
- Crosilla, F., Visintini, D., Sepic, F., 2005. A segmentation procedure of LIDAR data by applying mixed parametric and nonparametric models, in: *IAPRS&SIS*, XXXVI, 3/W19, Enschede, 132-137.
- Crosilla, F., Visintini, D., Sepic, F., 2007. An automatic classification and robust segmentation procedure of spatial objects. *Statistical Methods & Applications*, 15, 329-341.
- Do Carmo, M.P., 1976. *Differential geometry of curves and surfaces*, Prentice-Hill International, London.
- Garland, M., 1999. Multiresolution modeling: Survey & future opportunities", *Eurographics '99*, State of the Art Report, Sept.
- Haala, N.; Reulke, R.; Thies, M.; Aschoff, T., 2004. Combination of terrestrial laser scanning with high resolution panoramic images for investigations in forest applications and tree species recognition, in: *IAPRS&SIS*, XXXIV, 5/W16, Dresden.
- Ohio State University: *OSU MSU/WSU Range Image Database*, <http://sampl.ece.ohio-state.edu/data/3DDB/RID/index.htm>

ACKNOWLEDGEMENTS

This work was carried out within the research activities supported by the INTERREG IIIA Italy-Slovenia 2003-2006 project "Cadastral map updating and regional technical map integration for the GIS of the regional agencies by testing advanced and innovative survey techniques".

The authors thank Francesco Crosilla for providing some analytical computations at the beginning of the research.

NOVEL PLATFORM FOR TERRESTRIAL 3D MAPPING FROM FAST VEHICLES

A. Blug, C. Baulig, M. Dambacher, H. Wölfelschneider, H. Höfler

Fraunhofer Institute for Physical Measurement Techniques,
Heidenhofstraße 8, D-79110 Freiburg, Germany
Andreas.Blug@ipm.fraunhofer.de

KEY WORDS: Mobile, Mapping, Fusion, Laser scanning, Terrestrial, Close Range, City, Model

ABSTRACT:

We present the novel THelix 3D mapping system running on the LIMEZ III measurement train which is able to record clearance data and visual information about the surroundings of railway lines at velocities up to 100km/h. Due to a unique integration of complementary optical sensors and a sophisticated data management, the system provides comprehensive geometrical information about all trackside objects in both, global and local co-ordinates. A new and extremely fast close range laser scanning arrangement, stereo photogrammetry images, light sheet technology, GPS/INS and forward view laser scanning are combined to overcome the typical disadvantages of each single method. All geometric results are registered automatically to track centre, enabling the calculation of the closest profile within 0.1m in real time. Stereo photogrammetry complements the laser scanner point cloud for missed objects. The result is geometrical and texture information acquired from different perspectives.

The presented system is quite a universal close range 3D measurement system suitable for many mobile mapping applications. A similar system mounted on a car might be used for large area 3D mapping in roads without obstructing normal traffic. The results can be used to complement aerial city models (LoD 2) with realistic façade information in order to build large area LoD 3 models.

1. INTRODUCTION

1.1 Clearance Gauge Measurements

Clearance gauge measurement of railway lines is crucial for safe operation with regular trains and even more for oversized loadings. Even small objects extending into the allowed clearance and small displacements of the track may cause considerable damage. In any case disturbance of the regular service due to data recording should be avoided if possible. This can only be achieved if the measurement trains are able to speed up to at least 100 km/h so that they are allowed to enter even high speed lines. This application therefore requires gapless 3D measurement, documentation and visualization of the track ambience at high acquisition speed.

The THelix measurement system (THelix stands for travelling helix, describing the movements of the laser spots) was designed by Metronom Automation GmbH, Mainz, FTI Engineering Network GmbH, Dahlewitz, and Fraunhofer Institute for Physical Measurement Techniques for the Deutsche Bahn AG in order to generate a 3D and visual documentation of the complete German railway network of about 40,000km once a year. Measurement quantities are the clearance profile with reference towards track centre of all objects within a distance of $\pm 3m$ along the track (including signs) with an accuracy of $\pm 20mm$ (3σ). 3D data is recorded within $\pm 8m$, including distance, height, gauge and cant of the adjacent track. Additionally, gauge and cant of the track under test and vehicle position in global co-ordinates are acquired.

Several approaches for clearance measurement were published in the past (Blondeau, 1999; Meixner 2002; Cybernetix 2003; Blug 2004). All of them apply optical measurement principles, and optics still meets the users' demands best. However, none of the known and applicable measurement principles satisfies all the requirements of modern railway lines.



Fig. 1: 3D mapping system on the measurement train. Components: GPS/INS system (a, f), video and stereo photogrammetry system (b), twin head laser scanners (c, g), forward looking scanners (d, e), track reference (h).

1.2 Other Applications

Regarding its components, the system is very similar to other mobile mapping systems used on the road (Gräfe 2003; Haring 2005). It delivers globally and locally referenced geometric and texture information needed for applications such as city modelling, terrestrial navigation or disaster engineering, e.g. virtual fire fighting. For many applications, a vehicle velocity

up to 100 km/h during acquisition is highly advantageous, because it allows for cost-effective data acquisition without obstruction of the traffic flow. Differences lie in the measurement range and in the reference to the road instead to the rails and in the evaluation of the data.

2. SYSTEM DESIGN

Fig. 1 shows the system in front of the measurement train. All components are mounted on a single measurement frame making it easy to migrate this system, e.g. onto the rear side of a car. To meet the clearance profile requirements, state of the art stereo photogrammetry was combined with two extremely fast laser scanners based on the twin head geometry. Additionally, it comprises a GPS/INS system for global localization and a sheet of light rail reference system. The manual evaluation of the photogrammetric data is supported by two forward looking scanners. Some more details can be found in Höfler, 2006.

2.1 Track reference system

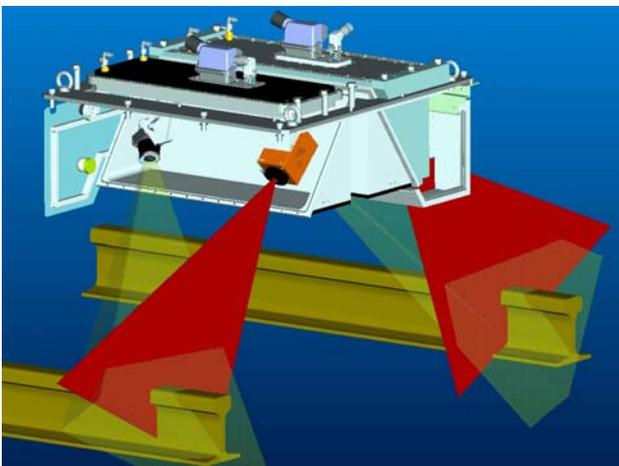


Fig. 2: Sheet-of-light track reference system

The track reference system (Fig. 2) consists of two sheet-of-light triangulation systems which are tilted 45° to measure horizontal and vertical position of both rails. The horizontal field of view is 0.56 m on each side. It measures rail position with a frequency of 50 Hz and an uncertainty of ± 1 mm for horizontal and vertical position. Therefore this system enhances the overall measurement uncertainty towards track centre of the laser scanners and the stereo photogrammetry system.

2.2 Twin Head Laser Scanners

Fig. 3 is a CAD drawing of a twin head laser scanner (Blug, 2004; Blug, 2005; Wölfelschneider 2005). Each scanner contains two fibre coupled measurement modules with sampling rates of 1 MHz for distance and 4 MHz for intensity. The distance is measured using the phase shifting technique (Wölfelschneider 2005) with two modulation frequencies resulting in unambiguous measurement ranges of about 1.2 and 20 meters. The laser beams of the two modules are directed on both surfaces of a double sided mirror and therefore propagating in opposite directions. The angle of the second scanner is fixed to 90° towards the first one. Each mirror rotates with 278 Hz. Therefore both scanners together acquire 1112 profiles per second with 3600 distance values and 14400 intensity values. These specifications allow for a minimum

object size of less than 30 mm in driving direction at a speed of 100 km/h.

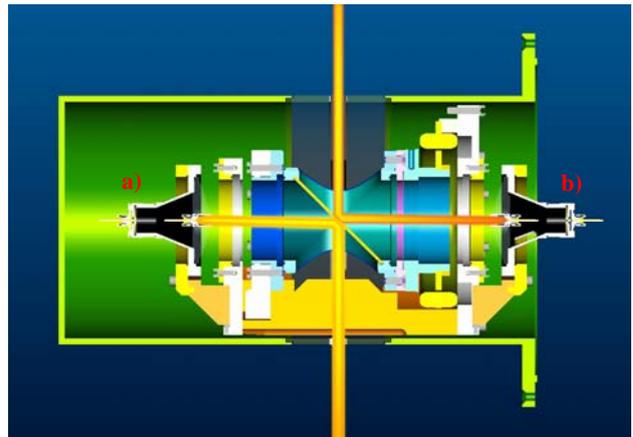


Fig. 3: Twin head laser scanner with two fibre coupled distance measurement modules (a, b) on a double sided mirror.

A very critical issue, especially for fast laser scanners, is the accuracy of measurement points near edges. There are three major limiting quantities: the spot size (which is 5 to 10 mm here, depending on distance), the dynamical response of the detector on changes in intensity and the response time - or bandwidth - of the distance measurement electronics. The bandwidth is independent from the sampling rate (here: bandwidth is 770 kHz, sampling rate 1 MHz) and a compromise between the noise and the response time of the system. Boehler published a target to assess the spatial resolution - which is similar to the edge quality - of laser scanners working in spherical co-ordinates (Fig. 5 in Boehler, 2003). This target requires an equal point density in scanning direction (here: vertical) and perpendicular to it (here: driving direction) which does not apply to the cylindrical scanning symmetry of our scanner. Therefore we constructed a target which combines changes in colour with spatial structures. It consists of several plates (width 5 cm) with different colours (black, grey, and white) mounted 0.6 meters in front of a black wall, half of the unambiguous measurement range of 1.2 meters. Therefore, it causes the maximum phase shift of 180 degrees for this modulation frequency.

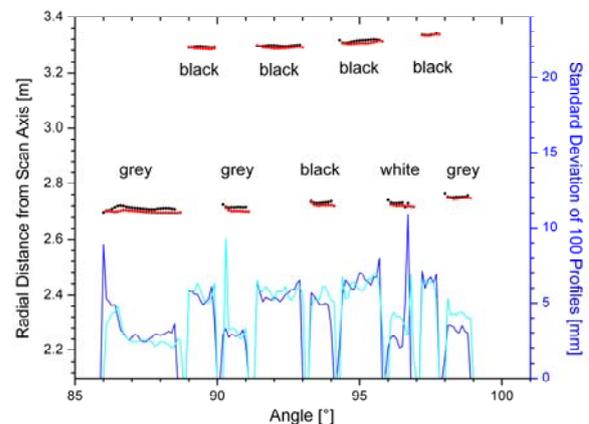


Fig. 4: Measurement accuracy of two distance measurement modules near edges.

Fig. 4 shows the results of the two modules obtained for that target in polar co-ordinates. The radii are averaged over 100 profiles. The standard deviation is given on the right side. A

standard deviation of zero means that those points were sorted out as uncertain by the point validation algorithm of the scanners. The graph shows systematic effects in the range of 5 to 10 mm an increased standard deviation near edges. The maximum number of disregarded neighbored points is two. This means, that edges can be measured with an accuracy of about ± 1 measurement point.

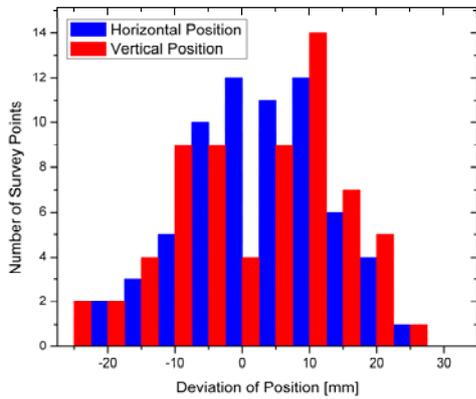


Fig. 5: Measurement deviations on survey points.

In Fig. 5, deviations (3σ) measured from the running train on 66 survey points are listed in the histogram. The survey points were the horizontal and vertical position of the edges of black and white cuboids mounted on masts towards track centre. At those positions the horizontal and the vertical position of the closest point of the track referenced point cloud was compared to tachymeter measurements (Leica TCR 407). Train velocity varied between 83 and 98 km/h and the smallest curve radius was 650 m. The radial measurement distance between survey points and laser scanners varied between 3 and 5 m. Therefore these deviations contain the uncertainties of single measurement points from the laser scanner point cloud, the track reference system and the reference measurements (estimated from control points to ± 5 mm, 3σ). As expected for laser scanners, the horizontal positions are slightly better (-18 to 25mm) than the vertical ones (-25 to 23 mm), due to the angular resolution of the scanner of 0.1° .

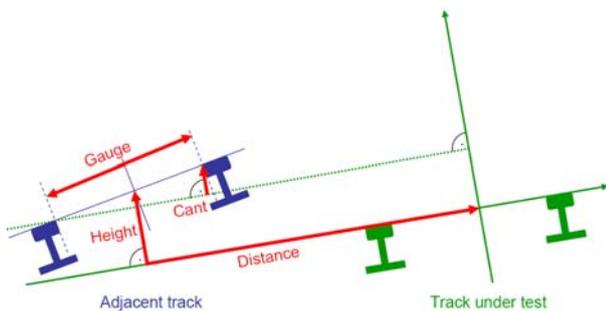


Fig. 6: Definition of adjacent track parameters.

The use of two scanners in a cylindrical symmetry also allows for the exploitation the intensity images for triangulation, which is necessary to achieve an accuracy of ± 8 mm (3σ) for the cant of the adjacent track (Fig. 6). For this purpose, the laser scanners are used as cameras with a focal depth defined by the width of the laser spots (Fig. 7).

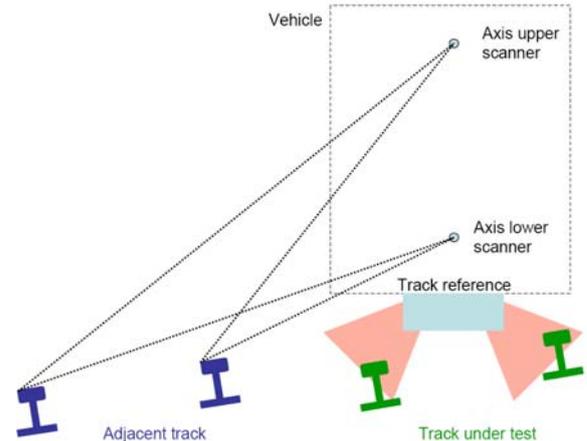


Fig. 7: Determination of the adjacent track parameters by triangulating the intensity information of the two scanners. The dotted lines indicate the angles of the rear rail edges.

For the adjacent track the intensity images of both scanners (Fig. 8) are evaluated for the angles of the rail edges by image processing whereas the underlying 3D data is used to identify the rails within the image. The red lines in Fig. 5 mark the rails identified in the underlying 3D data, the blue ones the rail edges determined in the intensity images with an angular resolution of 0.025° . It therefore improves the angular accuracy of the laser scanners.

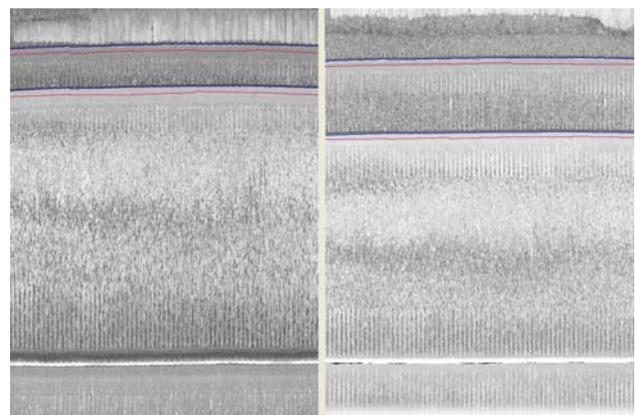


Fig. 8: Intensity images of the adjacent track from the lower scanner (left) and from the upper one (right).

2.3 Video and Stereo Photogrammetry System



Fig. 9: Carbon base and housing of the video photogrammetry system

The video photogrammetry system comprises four 2-megapixel monochrome video cameras, mounted on a stable, warp resistant and temperature inert carbon base in an air-conditioned housing on top of the mounting frame (Fig 9). They group into two pairs of stereo photogrammetry systems covering the left and the right part of the measurement range, respectively.

Fig 10 shows a schematic of the field of view of the different cameras. The measurement range amounts to 8 meters

horizontally, 5 meters vertically and 2 meters longitudinally. A set of pulsed infrared LED modules provide illumination of the measurement range and an intelligent brightness sensor, i.e. a fast intelligent CMOS-camera, controls the exposure time of the cameras according to the ambient light.

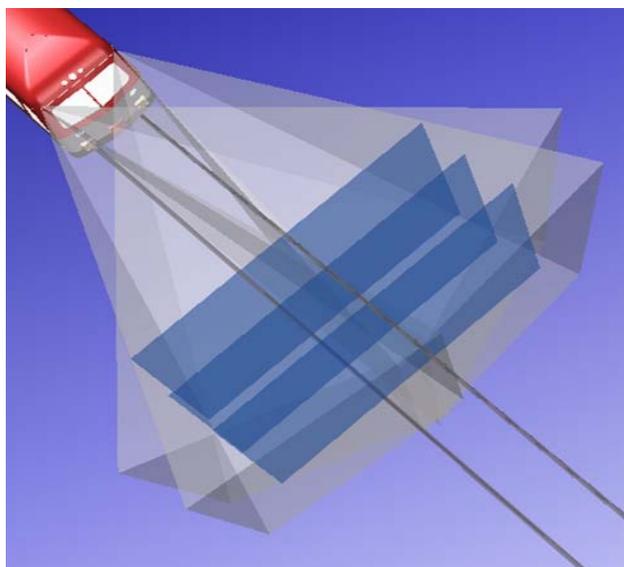


Fig. 10: Schematic diagram of the field of view of the two photogrammetry cameras

Triggered by a position encoder the cameras simultaneously accumulate pictures every 2 meters covering the whole measurement range along the track. At maximum speed this sums up to 14 sets of pictures per second. They are compressed and dumped to disk in real time. The measurement accuracy is currently under evaluation.

2.4 Forward view laser scanners

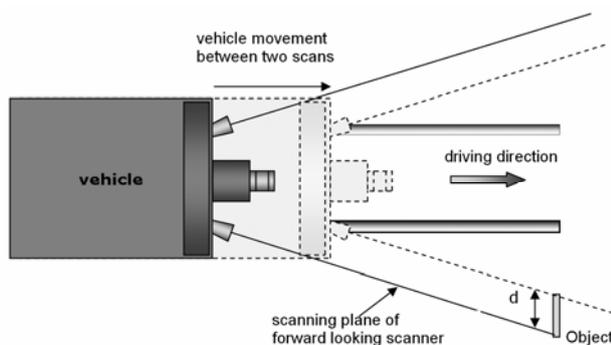


Fig. 11: Top view of the arrangement and the scanning plane of the forward looking scanners.

The video system together with the described side view laser scanners ensure that no trackside object is ignored. However, to identify all objects may be very time consuming if all videos have to be watched by an operator during offline data evaluation. There is still no image processing algorithm available allowing the reliable identification of any imaginable trackside object automatically. That is why two additional forward looking laser scanners are installed. Scanning plane is vertical with an angle of about 10° outwards of the driving direction. As sketched in Fig 11, the minimum detectable object size d is given by the viewing angle, the scanning frequency and the speed of the car. Assuming an angle of 10° , a speed of

100 km/h and a scanning frequency of 50Hz results in $d = 100$ mm. Therefore, these scanners are able to detect any objects with an extension down to 100 mm perpendicular to the track, independent of their thickness. Together with the side viewing scanner this configuration allows the detection of any trackside object and minimizes the time for the offline evaluation of the video sequences.

2.5 GPS/INS system

The system is equipped with most modern GPS and INS positioning systems using laser gyroscope, accelerometers and differential GPS. The post processing using offline reference data leads to a location precision in the few centimetres range.

3. DATA PROCESSING

3.1 Real Time Data Processing

All laser scanners and the track referencing system are evaluated automatically and in real time in order to minimize post processing time and the amount of stored data. It also enables online visualisation of the recorded data so that the operator is able to judge the measurement quality during the measurement run. The following steps are carried out:

1. All systems with exception of the GPS/INS align their data to a pulse coming every 0.1 meters of travelling distance.
2. The track reference system extracts the position of the rails in vehicle co-ordinates.
3. Both twin head laser scanners calculate the narrowest profile within 0.1 meters.
4. Registration: Transformation of the narrowest profile from both scanners to track co-ordinates.
5. The images of both twin head scanners (Fig. 8) are combined and the adjacent track is extracted. Only adjacent track parameters are stored.
6. The forward looking scanners calculate the narrowest profiles and count infringements into a predefined reference profile.
7. Visualization of the 3D point cloud, track parameters, adjacent track position, and infringements from the forward scanners.

The point density of 0.1 meters in driving direction is sufficient for visualization of the point cloud. The calculation of the narrowest profile in step 3 assures, that the relevant points of smaller objects (down to 30 mm in driving direction) are not lost. At the same time, the amount of data from the twin head scanners is reduced from 70 GB per hour to distance dependant amount between 100 MB per km in tunnels and down to 30 MB per km on free tracks where large parts of the scan profiles do not contain targets within the measurement range. Table 1 lists the amount of data stored for a measurement distance of 500 km which can be acquired within one day.

Measurement system	Amount of data [GB]
Twin head laser scanners	50,3
Video documentation	33,6
Stereo photogrammetry	178,9
GPS/INS (time dependant, 10h)	1,1
Total data amount	263,9

Table 1. Data amount of a 500 km measuring run

3.2 Post Processing

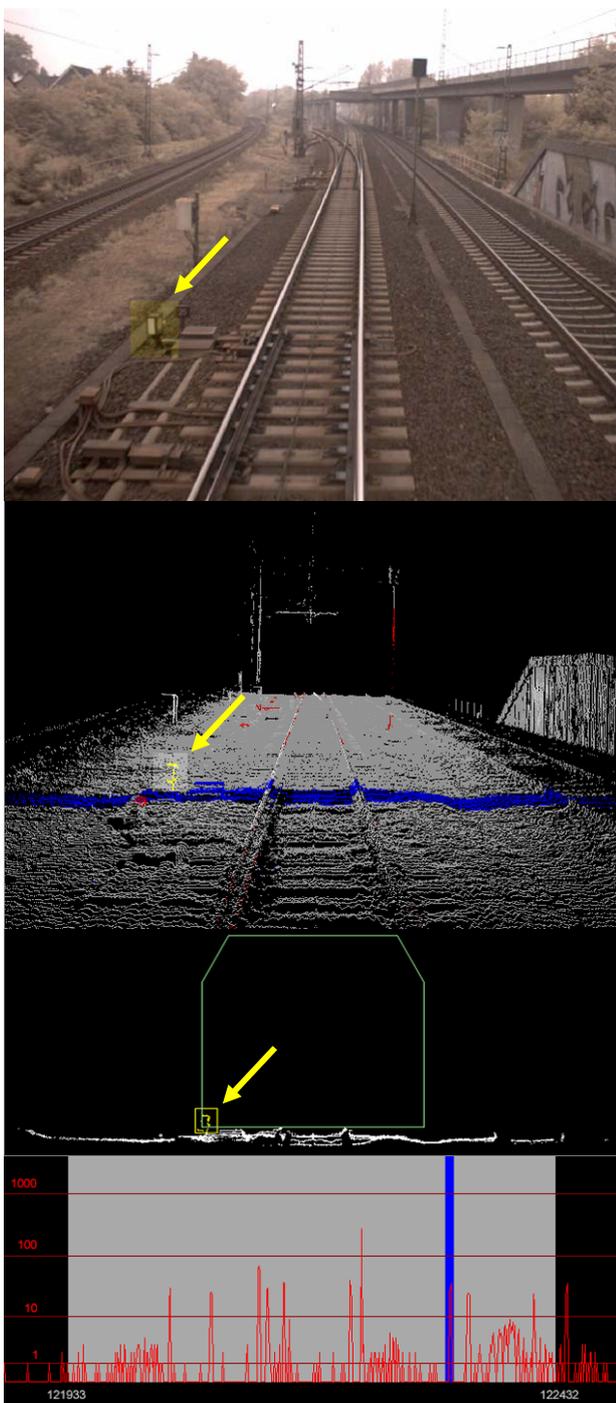


Fig. 12: Video documentation image (top); 3D point cloud from 500 profiles or 50 m (second); 2D view of the blue profiles from the 3D point cloud (third); infringement statistics with the number of points inside of a reference profile (bottom). The red points of the point cloud lie inside the reference profile. The object under test is marked by yellow arrows.

The real time data alignment and registration reduces post processing to two steps:

1. Refinement of the GPS data by differential reference data (ASCOS).
2. Selection and documentation of objects infringing a predefined reference profile.

For the object selection, the 3D point cloud is combined with visual information from the documentation video and with an infringement statistics showing the number of points within a reference profile over track distance (Fig. 11). The object contour is selected in a 2D view. The comparison between video and 3D point cloud visualizes objects missed by the laser scanners, which are typically signs. Their edges are determined by stereo photogrammetry.

4. CONCLUSIONS

The system exploits the complementary properties of stereo photogrammetry and laser scanning. Laser scanning allows for automatic referencing and evaluation of 3D data. The results are verified by the video images and remaining gaps are evaluated manually using stereo photogrammetry. Both together cover almost the complete surroundings at travelling velocities up to 100 km/h and allow for efficient post processing. The total data volume of all systems is about 260 GB for 500 km.

The main purpose of the LIMEZ III train is clearance gauge measurement. Nevertheless, the THelix measurement system provides gapless geometrical and texture information useful for many other applications. In particular, mobile mapping applications profit from the high acquisition speed, because the obstruction of traffic is minimized. Therefore a similar sensor arrangement mounted on a car might be a promising solution to acquire realistic façade information for large area city models on a LoD 3 level. There, the high point density of the laser scanners is advantageous for an automated evaluation like the registration of point clouds and images (Boehm, 2007).

5. REFERENCES

- Blondeau, J., 1999. Limez II der Lichtraummesszug der Deutschen Bahn. *Der Eisenbahningenieur*, 50 (1999) 2, pp. 58-64
- Boehler, W., 2003. Investigating Laser Scanner Accuracy. XIXth CIPA Symposium, Antalya, Turkey, October 2003
- Boehm, J., Becker, S., 2007. Automatic Marker-free Registration of Terrestrial Laser Scans using Reflectance Features. *Proceedings Optical 3-D Measurement Techniques VIII, Volume I*, pp. 338, Zürich, Switzerland, July 9-12, 2007, ISBN 3-906467-67-8
- Blug, A., 2004. Fast Fiber Coupled Clearance Profile Scanner using Real Time 3D Data Processing with Automatic Rail Detection. *Proceedings IEEE Intelligent Vehicles Symposium*, pp. 658-663, Parma, Italy, June 14-17, 2004
- Blug, A., 2005. Novel Geometry for a High Speed Outdoor Laser Scanning System. *Proceedings Optical 3-D Measurement Techniques VII, Volume I*, pp. 154, Vienna, Austria, October, 3-5, 2005, ISBN 3-9501492-2-8
- Cybernetix, 2003. Observer. Unattended high speed clearance measurement. Brochure by Cybernetix, Marseille, France version 1.2, 2003. <http://www.cybernetix.fr>
- Gräfe, G., 2003. Mobile Mapping with Laser Scanners using the MOSES. *Proceedings Optical 3-D Measurement Techniques VI, Volume I*, pp. 381-388, Zürich, 2003. ISBN 3-906467-43-0.
- Haring, A., 2005. Vehicle-Born Acquisition of Urban Street Space by Combined Use of 3D Laser Scanning and Close-

Range Photogrammetry. Proceedings Optical 3-D Measurement Techniques VII, Volume I, pp. 275-284, Vienna, 2005. ISBN 3-9501492-2-8.

Höfler, H., 2006. High Speed Clearance Profiling with Integrated Sensors. Proceedings 7th World Congress on Railway Research, Montreal, Canada, June 4-8, 2006

Meixner, J., 2002. Bildgebendes Laserradar Präzisionsvermessungssystem. EI-Eisenbahningenieur (53), 5/2002, pp. 32.

Wölfelschneider, H., 2005. Fast Distance Measurement for Laser Scanning, tm Technisches Messen 72 (2005) 8

AN INTEGRATED APPROACH TO LEVEL-OF-DETAIL BUILDING EXTRACTION AND MODELLING USING AIRBORNE LIDAR AND OPTICAL IMAGERY

F. O. Chikomo*, J. P. Mills, S. L. Barr

School of Civil Engineering and Geosciences, University of Newcastle upon Tyne, Cassie Building, Newcastle upon Tyne, NE1 7RU, UK – (F.O.Chikomo, J.P.Mills, S.L.Barr)@ncl.ac.uk

WG III/4 - Automatic Image Interpretation for City-Modelling

KEY WORDS: Buildings, 3-D, Reconstruction, Level-of-Detail, Laser scanning, Imagery, Integration

ABSTRACT:

The need to automatically extract topographic objects, especially buildings, from digital aerial imagery or laser range data remains an important research priority in both photogrammetric and computer vision communities. This paper describes the proposed model for Level-of-Detail building modelling and progress with the prototype implementation. The paper begins with an overview of the concept of Level-of-Detail, important for adaptive building modelling. Building regions of interest are derived from a normalised digital surface model (nDSM) and regularisation of the roof lines is achieved by a set of contextual constraints with particular emphasis on rectangular buildings. For detailed building reconstruction, the main consideration is given to polyhedral building types with limited support for curvilinear shapes. A moving least squares approach for computation of surface normal vectors and texture metrics is employed for planar segmentation of both gridded data and unstructured point clouds. Delineation of homogeneous planar segments is based on a distance metric between neighbouring local planes. 2-D edge lines derived from the orthoimage are matched with 3-D lines derived from LiDAR based on adjacent plane intersections and then used for the final building reconstruction. Connected regions which fail the local planarity tests and are sufficiently large, are segmented using curvature measures based on least squares quadric surface fitting. Provisional results from the algorithms are promising.

1. INTRODUCTION

1.1 Background

The need to automatically extract 3-D building data from digital aerial imagery or laser range data remains an important research priority in both photogrammetry and computer vision. 3-D building data is important for applications such as city modelling, environmental engineering, disaster mitigation and management and emerging civilian and tactical applications such as virtual and augmented reality and homeland security. Significant success has been achieved so far with semi-automatic building extraction systems using either imagery or LiDAR data, however in restricted domains. The need for an integrated data approach for building extraction has been realised however not yet fully tested. To meet the varying demands in terms of capture of building detail and representation, incorporation of Level-of-Detail (LoD) mechanisms into the building extraction schema has become a necessity. This paper presents an integrated approach for Level of Detail building model reconstruction using airborne LiDAR data and optical imagery and discusses the prototype implementation of the proposed model.

1.2 Related Work

Several algorithms have been proposed for automating the three-dimensional reconstruction of buildings however a robust and versatile solution is yet to be found although significant progress has been made. A discussion of systems based on and accuracies obtainable with photogrammetry and laser scanning in building extraction is contained in Kaartinen et al. (2005). To date, building extraction has largely been based on single data sources, in most cases either LiDAR data alone (Vosselman, 1999; Verma et al., 2006) or images alone (Scholze et al., 2001;

Kim and Nevatia, 2004) however the current trend is on integrated data paradigms. Integration of data sets provides multiple cues that can ease the problem of building reconstruction and result in significantly higher levels of automation in the algorithms. The data bases for integrated approaches have included multiple geometric data, GIS layers and bespoke or scene specific knowledge. A number of researchers have demonstrated approaches for combining data for building modelling, for example LiDAR and aerial images (Rottensteiner et al., 2004), LiDAR and three-line-scanner imagery (Nakagawa and Shibasaki, 2003), LiDAR and high resolution satellite images (Sohn and Dowman, 2001), LiDAR and 2-D maps (Overby et al., 2004), aerial images and 2-D maps (Suveg and Vosselman, 2004) and LiDAR, 2-D maps and aerial images (Vosselman, 2002). Schenk and Csatho (2002) discuss theoretical frameworks for multi-sensor data fusion for generic surface description. Not many researchers however have incorporated LoD modelling into their approaches, important for catering for diverse users and applications.

1.3 Level-of-Detail Modelling

The concept of level of detail has been used in computer graphics since the 1970s, mainly for increasing the efficiency of object rendering. Rendering efficiency is achieved by decreasing the visual and geometric detail of 3-D objects as a function of distance from the view point or other metrics such as the perceived object importance. The concept has been adapted and extended for city modelling by the Special Interest Group on 3-D Modelling (Sig3D of the GDI Initiative). LoD building modelling involves adaptable and scaleable extraction and representation of building model information. This enables the capture of building information to be varied depending on the specific requirements of the project and limitations of the building extraction techniques and input data characteristics.

* Corresponding author.

The five levels of detail range from a simple block (2.5-D) model right up to a walkable model, which takes into account both internal and external geometric detail. Figure 1 below illustrates the concept of level of detail modelling as modified for this research. For this research, the aim is to work up to LoD2 with texture effects applied. LoD modelling is important for understanding the trade-offs between model detail and automation potential.

One can possibly identify two main approaches for incorporating the LoD schema into the building reconstruction process. Firstly, a bottom-up approach where a multi-level strategy is adopted for reconstructing each level more or less independently varying the data sets used, their resolution and the algorithmic detail. Secondly, the initial reconstruction effort might be aimed at a detailed level with lower levels derived by building generalisation. A hybrid approach is also possible. In this research, a bottom-up approach is adopted for the multi-level building reconstruction.

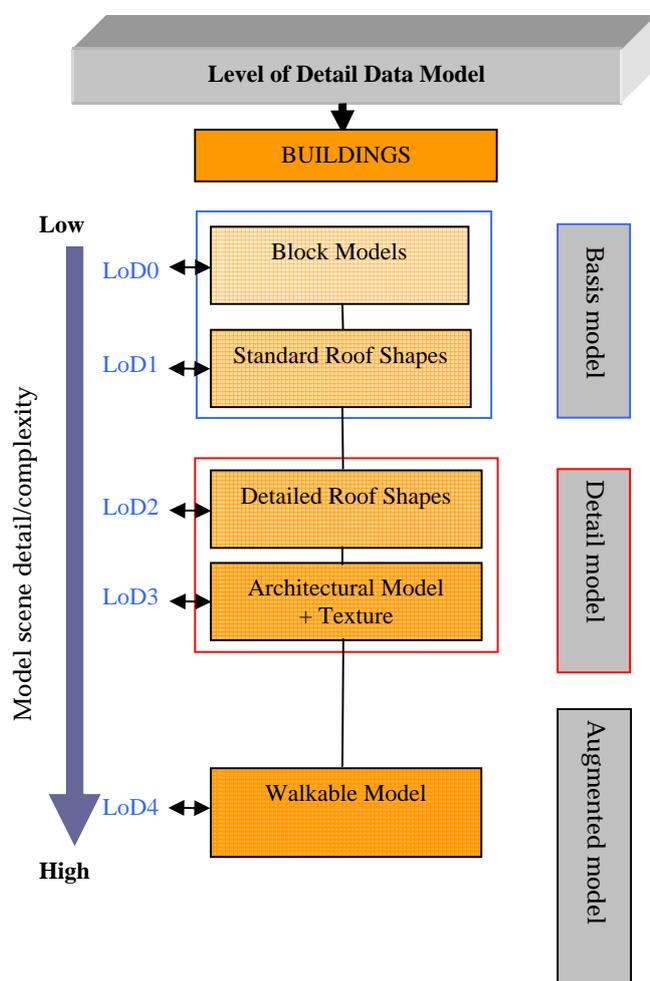


Figure 1. Level-of-Detail modelling schema

1.4 Study Area and Datasets

The study area for the research is Portbury, a small agricultural village, approximately 11 kilometres North West of Bristol, England. Portbury is one of the test sites identified by the Ordnance Survey Research Labs, source of LiDAR data and digital imagery, for research on automated building extraction. The data sets available for the test site are as follows: LiDAR

data [16 points/m²]; digital orthoimagery [GSD 10cm] and OS MasterMap Topography data.

2. ROOFLINE DETECTION

The rooflines are detected from a normalised DSM (nDSM) obtained by differencing an optimised DTM derived from LiDAR using the adaptive TIN algorithm employed in the commercial software TerraScan and an interpolated DSM. A three-metre threshold is applied to the nDSM to detect above ground objects however these include buildings and vegetation. The options followed for vegetation removal included use of intensity data, generic tree point classification, least squares planar fitting differences and near infrared image analysis. The minimum building size was considered to be 12m² according to Ordnance Survey specifications. Building segment simplification is achieved using the modified sleeve algorithm and rectangular enforcement is achieved by deriving a moments based orientation and enforcing building line segments to be perpendicular or parallel to this orientation within a defined tolerance. The extracted rooflines define the regions of interest for the geometrical reconstruction of the different levels of detail mentioned before.

3. BUILDING RECONSTRUCTION

3.1 LoD Reconstruction Schema

Reconstruction of LoD0 (block models) requires building rooflines and representative building heights. The building heights are derived from differences of roof and ground heights obtained from a DSM/DTM. LoD1 requires rooflines and LiDAR data for generic roof modelling. LoD2 additionally requires high resolution image data for accurate edge and sub-feature detection and texturing the models.

3.2 Types of Building Surfaces

For this research, consideration is given to two types of surfaces, planar and curvilinear surfaces as shown in Figures 2 and 3 respectively. Planar surfaces are by far the most common type and form the initial hypothesis for the reconstruction algorithms. For curvilinear surfaces, consideration is given to quadric surfaces although this could be extended to more generic superquadric surfaces.

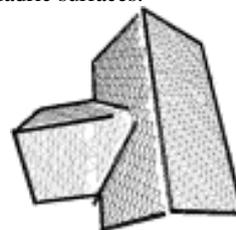


Figure 2. Planar surfaces

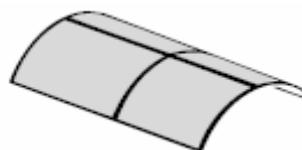


Figure 3. Quadric (curvilinear) surface

3.3 Planar segmentation

For planar segmentation of the LiDAR data, a moving least squares plane fitting algorithm is employed. For each point, a cluster of neighbouring points is determined depending on whether a grid or an unstructured point cloud is used. For gridded data, the algorithm searches for the 8-connected neighbours of each point however the grid resolution can be changed. For unstructured point clouds, the neighbourhood of each point is defined based on either a search radius or a defined number of nearest neighbours within a set maximum distance.

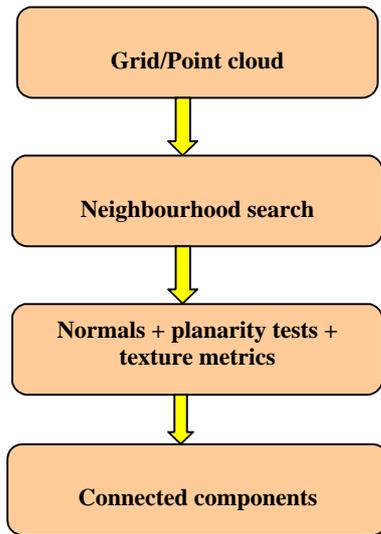


Figure 4. Workflow for planar segmentation

A least squares plane is fitted to the neighbourhood of each point. For each point, a normal vector and texture metrics are computed together with tests for local planarity. Clusters of points lying on the same planes are determined by comparing similarity of normal vector orientations and distances between their local planes. Figure 4 illustrates diagrammatically the steps in the planar segmentation phase of the algorithm. Figure 5 illustrates the vector dispersion and computation of a normal vector for a planar patch. The requirement for this algorithm is to work with both airborne and terrestrial LiDAR data in order to meet the requirements of the different levels of detail.

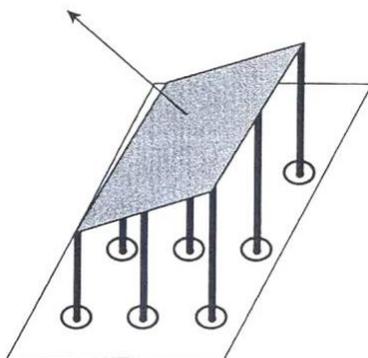


Figure 5. Planar patch normal vector (Parker, 1996)

Applying the planar segmentation to gridded data makes the neighbourhood search easier and allows other image based metrics such as texture coefficients to be calculated however interpolating the point data introduces some unwanted artefacts. Working on scattered point data introduces a computational

overhead and requires appropriate adaptation for image based metrics. We experimented with a search radius of 1m. The planar segmentation algorithm can be summarised as follows:

1. For each data point, locate corresponding points falling within the defined neighbourhood of the point based on the appropriate criteria.
2. For each defined neighbourhood, compute coefficients of the plane and from that compute the normal to the plane for that neighbourhood and the azimuth of the projection of the normal on the x,y plane.
3. Compute metrics for assessing local planarity (section 3.5) together with a texture descriptor for each neighbourhood and numerical checks for validating the least squares computation.
4. Data points assumed locally planar are further clustered into consistent planar regions.

3.4 Localised planar fitting of 3-D points

Given a defined neighbourhood $\{(x_i, y_i, z_i)_{i=1,m}\}$ of a point, we want to fit a plane which satisfies the relationship:

$$Z = Ax + By + C \tag{1}$$

where Z represents height, A and B are slope parameters in the x and y direction respectively and C is the offset at the origin.

In the implementation of the algorithms, we consider two forms of minimisation of the residual sums. The first form minimises the sum of residuals in the z -direction and the other considers the errors measured orthogonal to the plane and requires use of eigensystem solvers.

3.5 Tests for local planarity

A number of metrics are computed during the point classification phase in order to check if the point is locally planar and weed out erroneous points. The numerical tests are as follows:

- Centre point residual

For each point under consideration, the difference between the actual height value and the height computed using the determined plane parameters should be below a set threshold (Figure 6). In our case, we set the threshold to 0.1 metres based on the average error computed using photogrammetric control points. Equations 2 and 3 mathematically describe the computation of plane coefficients and the centre point residual respectively.

P_1	P_2	P_3
P_4	P_5	P_6
P_7	P_8	P_9

Figure 6. Centre point (p_5) and connected neighbours

$$\left. \begin{aligned} P_i &= (x_i, y_i, z_i) \forall i = 1..9 \\ E(A, B, C) &= \sum_{i=1}^9 [(Ax_i + By_i + C) - Z_i]^2 \\ \nabla E &= (0,0,0) \end{aligned} \right\} \tag{2}$$

$$p_5(\text{actual} - \text{calculated}) \leq \text{threshold} \tag{3}$$

- Eigen-analysis

The smallest eigenvalue of the dispersion matrix A (Equation 4) computed by reducing neighbourhood points to centre and taking product sums expresses the deviation of the points from the fitted plane. Local planarity is assumed if the smallest eigenvalue is less than a set threshold.

$$A = (P - M)^T (P - M) \quad (4)$$

where P is a matrix of neighbourhood points (x_i, y_i, z_i) and M the mean matrix.

A threshold of 0.05 metres was used for the eigen-analysis. This measure of planarity works well for gridded data however a more useful and standardised measure for planarity testing is the ratio of the smallest eigenvalue to the total variance. Fransens (1996) employs a similar method of comparing eigenvalues of the covariance matrix for planarity testing of data in an octree.

- Residual norms

For the residual vectors, the 2-norm (equation 5) is computed and gives a measure of the quality of fit of the model to the data points.

$$\|v\|_2 = (v^T v)^{\frac{1}{2}} \quad (5)$$

3.6 Connected components analysis

The next step is to determine if neighbouring sub-planes defined at each data point could lie on the same planar surface. Our first approach for clustering coplanar points was based on analysing the histograms of normal orientations for each building region of interest and using local peaks for grouping. This approach works well for sloped roofs and fails for flat roofs where the normal vector orientations can shift full circle. A more effective metric for coplanarity considers the distance between neighbouring sub-planes. For each point, the maximum distance between the sub-plane under consideration and surrounding locally planar neighbours. Points with plane distances below a defined threshold are then grouped into consistent clusters. The convex hull of each cluster is then extracted to define boundaries between planes. Final plane parameters for each cluster are determined using the RANSAC algorithm for robustness. The RANSAC algorithm ensures robust fitting of models in the presence of data outliers and requires a large sample of data points. To ensure convergence of the RANSAC algorithm, the inlier threshold is set to 0.05 metres, inlier percentage 75% and the maximum number of iterations 20.

3.7 Planar adjacency and 3-D lines

Planar adjacency graphs are determined for clusters found for each building region of interest. Adjacency is based on the distance between the outer boundaries of the clusters.

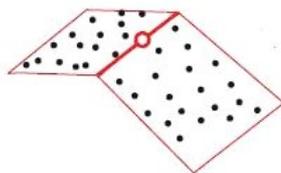


Figure 7. 3-D breakline determination (adapted and modified from Briese, 2004)

Adjacent planes are then intersected to determine 3-D breaklines from LiDAR data, which are then verified and

matched against edge lines from the orthoimage. Figure 7 illustrates the basic concept of a 3-D line description derived by intersecting planar patch pairs determined on the basis of clustered point cloud data with the circle representing the support point. The handling of step edges requires further consideration.

3.8 Quadric segmentation

Connected points that are not locally planar and form a sufficiently large size more than $1m^2$ are further tested for curvilinearity. A second degree polynomial surface is fit to the data points and takes the form:

$$Z = a_{20}x^2 + a_{02}y^2 + a_{11}xy + a_{10}x + a_{01}y + a_{00} \quad (6)$$

The steps in the segmentation can similarly be summarised as follows:

1. For each locally non-planar point, define a sufficiently large neighbourhood of points.
2. Use least squares to fit a quadric surface to the local neighbourhood of each point.
3. Compute the derivatives of the surface and the slope of the tangent at each point.
4. Compute curvature measures (Gaussian, Mean and Laplacian) using the derivative and slope of tangent.

A scale factor is applied for numerical optimisation of the computation. Surface modelling is considered additive and applied in a Constructive Solid Geometry fashion. The model could be extended to superquadric surface fitting for more generic curvilinear modelling.

4. IMAGE DATA ANALYSIS

Image data serves to provide more accurate breaklines for detailed modelling, verify 3-D lines derived from LiDAR and allow texture mapping for photorealistic building modelling. There are three important considerations for the integration of high resolution image data:

- Building localisation
Building roofline polygons are localised in the orthoimage by dilating the minimum bounding rectangle then projecting this into the image. A factor of 1.25 is applied to the areal dimensions of the minimum bounding rectangle. This reduces the search space for matching purposes.
- Edge extraction
2-D linear segments are extracted from the orthoimage using the Canny operator implemented in the open source computer vision library, OpenCV.
- Image pruning
The use of high resolution images (GSD 10cm) is required for higher levels of detail. A threshold is applied to remove spurious and short linear segments. To allow matching of the roofline polygons, linear segments parallel to the rooflines, within some tolerance, are retained for shape matching.

5. Provisional results and discussion

This paper outlined a methodology for level of detail building model reconstruction following an integrated data paradigm. A prototype implementation of the proposed model is in progress however most parts of the algorithm have been tested piecemeal. Figure 8 shows an orthoimage of the Portbury test with an area of interest linked to Figures 9 and 10 shown in red.

The test site contains mostly polyhedral buildings. Reconstruction of block models (LoD0) is achieved with minimal effort after processing the DSM/DTM. The planar segmentation algorithm (LoD1+) was applied to interpolated data at four grid resolutions, 1m, 0.5m, 0.25m and 0.10m in order to assess model sensitivity and also applied to the original point clouds.



Figure 8. Orthoimage of the test site (Portbury, Bristol, UK), in red an area of interest linked to Figures 9 and 10

Figures 9 and 10 show colour coded normal vector orientation maps of the small area of interest (in red, Figure 8), derived from the least squares planar fitting algorithm applied over grid DSM data at 25cm and 10cm resolutions respectively. The grid data was derived from point data at 16 points/m². 25cm resolution results in a more natural look whereas finer detail is apparent at 10cm resolution. The maps highlight planar surfaces, buildings in particular, very clearly. The reconstruction algorithms are applied to the extracted rooflines. Figure 11 shows an enlarged part of the study site and Figure 12, a colour coded normal vector map in the building regions of interest at 0.50m resolution. The map shows an efficient characterisation of building plane surfaces. The choice of grid resolution needs to be optimised in order to average out noise effects, retain required building geometric detail and optimise the turnaround time.

The planar segmentation algorithm could in principle be applied to terrestrial LiDAR data to meet requirements of LoD3 and above in our modified schema (described in Figure 1).



Figure 9. Colour coded normal vector map of the area of interest shown in Figure 8 derived from a gridded LiDAR DSM (25cm resolution) over the area

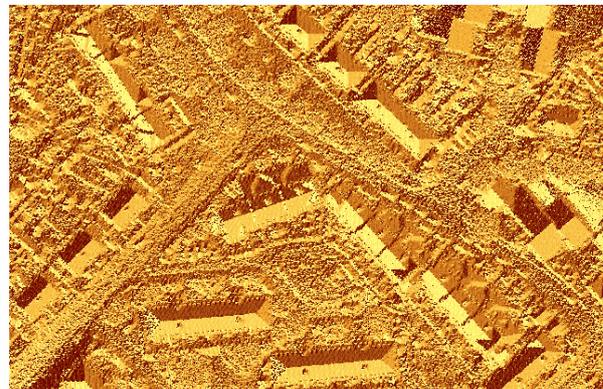


Figure 10. Colour coded normal vector map of the area of interest shown in Figure 8 derived from a gridded LiDAR DSM (10cm resolution) over the area



Figure 11. An enlarged part of the orthoimage for the test site showing building regions of interest linked to Figure 12, different from that shown in Figure 8

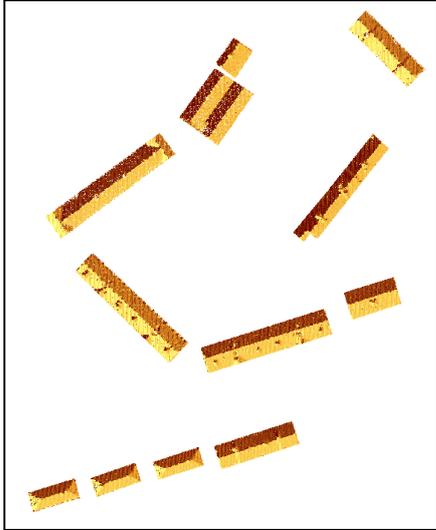


Figure 12. Colour coded normal vector map corresponding to image subset in Figure 11

The curvilinear segmentation algorithms have been tested over a different study area with appropriate building types. A significant level of automation in the planar and curvilinear segmentation has been possible. Further work is on labelling connected components and incorporating image edge lines into the segmentation and model reconstruction process. The results are promising.

6. REFERENCES

- Baltsavias, E., Greun, A. and Van Gool, L. (2001). Automatic Extraction of Man-Made Objects from Aerial and Space Images. Balkema Publishers, Rotterdam.
- Briese, C. (2004). Breakline Modelling from Airborne Laser Scanner Data. Institute of Photogrammetry and Remote Sensing. Vienna, Austria, Vienna University of Technology: 74pp.
- Chen, L. C., Teo, T. A., Shao, Y. C., Lai, Y. C., Rau, J. Y. (2004). Fusion of LiDAR data and optical imagery for building modelling. ISPRS Commission IV Workshop 7, Istanbul.
- Chikomo, F. O., Mills, J. P., Barr, S. L. (2006). Experimental comparison of LiDAR and image based semi-automatic building extraction systems. Paper presented at ILMF 2006, Denver, Colorado, USA.
- Fransens, J. (2006). Statistical segmentation for computer graphics. Expertise Centre for Digital Media. Hasselt, University of Hasselt, Belgium. PhD Thesis: 124pp.
- Kaartinen, H., Hyypää, J., Gülch, E., Hyypää, H., Matikainen, L., Vosselman, G., Hofmann, A.D., Mäder, U., Persson, Å., Söderman, U., Elmqwist, M. Ruiz, A., Dragoja, M., Flamanc, D., Maillet G., Kersten, T., Carl, J., Hau, R., Wild, E., Frederiksen, L., Holmgaard, J. and Vester, K. (2005). EuroSDR building extraction comparison. ISPRS Hannover Workshop 2005 "High-Resolution Earth Imaging for Geospatial Information". May 17-20, 2005, Hannover, Germany.
- Kim, Z., Nevatia, R. (2004). Automatic description of complex buildings from multiple images. Computer Vision and Image Understanding 96(1): 60-95.
- Nakagawa, M., Shibasaki, R. (2003). Integrating high resolution airborne linear CCD (TLS) imagery and LiDAR data. 2nd GRSS/ISPRS Joint. Berlin, Spring Verlag.
- Overby, J., Bobum, L., Kjens, E., Ilsoe, P. M. (2004). Automatic 3D building reconstruction from airborne laser scanning and cadastral data using Hough Transform. ISPRS Commission III: 6pp.
- Parker, J. R. (1996). Algorithms for image processing and computer vision. United States of America, John Wiley & Sons, Inc. 417pp.
- Paternell, M., Steiner, T. (2004) Reconstruction of piecewise planar objects from point clouds. Computer-Aided Design 36: 333 - 342.
- Rottensteiner, F., Trinder, J., Clode, S., Kubik, K., Lovell, B. (2004). Building detection by Dempster-Shafer Fusion of LIDAR Data and Multispectral Aerial Imagery. ICPR'04.
- Schenk, T., Csatho, B. (2002). Fusion of LIDAR Data and Aerial Imagery for a More Complete Surface Description. ISPRS Commission III, Symposium 2002: 8pp.
- Scholze, S., Moons, T. and Van Gool, L., 2001. A probabilistic approach to roof patch extraction and reconstruction. In: [Baltsavias et al., 2001]
- Sohn, G., Dowman, I. (2001). Building extraction using LiDAR DEMs and IKONOS images. ISPRS Commission III, WG III/3: 7pp.
- Stamos, I., Allen, P. K. (2000). 3-D Model Construction Using Range and Image Data. Computer Vision and Pattern Recognition_2000: 8 pp.
- Suveg, I., Vosselman, G. (2004). Reconstruction of 3D buildings from aerial images and maps. ISPRS Journal of Photogrammetry & RS 58: 202-224.
- Verma, V., Kumar, R., Hsu, S. (2006). 3D building detection and modelling from aerial LiDAR data. 2006 IEEE Computer Society Conference Computer Vision and Pattern Recognition 2: 2213 - 2220.
- Vosselman, G. (1999). Building Reconstruction Using Planar Faces in Very High Density Height Data. International Archives of Photogrammetry and Remote Sensing 32, part 3/2W5: 87-92.
- Vosselman, G. (2002). Fusion of laser scanning data, maps and aerial photographs for building reconstruction. IGARSS'02, Toronto, Canada.

7. ACKNOWLEDGEMENTS

We would like to thank OS Research Labs for providing data sets and steering the direction of this research. We also thank NIIRS10 and GTA Geoinformatik GmbH for their assistance with other aspects of the research.

A SEMI-AUTOMATIC HIGH RESOLUTION SAR DATA INTERPRETATION PROCEDURE

F. Dell'Acqua, P. Gamba, G. Lisini

Remote Sensing Group, department of Electronics, University of Pavia, via A. Ferrata, 1, I-27100 Pavia, Italy,
{name.surname}@unipv.it

Commission III, WG III/5

KEY WORDS: SAR, scene interpretation, object-based analysis

ABSTRACT:

This paper describes a semi-automatic procedure for cartographic mapping using high resolution SAR data. Two-dimensional and three-dimensional feature extractions are combined in order to achieve a basic yet effective recognition of the elements in the scene. Many interesting elements of the landscape are automatically extracted without requiring a large interaction with the operator. In particular, the procedure is well suited for detecting man-made features, such as the road network (outside and inside human settlements) and built up areas. It can be used, however, to extract higher level elements of the scene, such as crossroads, bridges and overpasses, by data fusion at the feature level of the previous extraction, because it is characterized by a multi-scale object-based approach.

1. INTRODUCTION

Land cover mapping (and consequently land use mapping) relies, at the level of detail required by most planners, on high spatial resolution sensors. Cartographic feature extraction has been traditionally obtained by optical and near infrared sensors, installed on airborne or, more recently, spaceborne platforms. Instead, due to the rather coarse resolution and the complexity of their data sets, radar sensors are usually not considered for these tasks. However, because of their interferometric capabilities, they are very useful and indeed they are almost universally exploited for their three-dimensional characterization of the landscape, i.e. for a different kind of cartographic feature extraction.

The availability or the development of new airborne and satellite SAR sensors, with high spatial resolution, and short revisit time, is going to change this "status quo". In particular, high spatial resolution SAR sensors are starting to provide enough information for the recognition and characterization of two-dimensional cartographic features within a radar image. This possibility, further enhanced by the straightforward combination of 2D and 3D information for sensors with interferometric capabilities, makes modern SAR system really appealing for mapping purposes.

This is testified by some recent papers using SAR for mapping of human settlements, down to the scale of single urban features. Examples are those presented in special sessions devoted to SAR urban mapping in the last IEEE Geoscience and Remote Sensing Symposium (IGARSS'06), European SAR conference (EUSAR06) and Joint Urban Remote Sensing Event (JURSE2007). These sessions highlights the fact that some urban features, which are among the most complex to be identified in remotely sensed images, can be obtained using SAR data with a satisfying accuracy level. No doubt that current optical data sets provide better input to cartographic practice, mostly because it is manually done, and radar data is very difficult to be analyzed by untrained (and sometimes also for trained) analysts. The point is that there are already available semi-automatic procedures that could provide outputs at a commercially effective level. However, they are often dedicated

to the extraction of single features, such as urban extent (He *et al.*, 2006), water bodies (Hall *et al.*, 2005), vegetation (Askne *et al.*, 2003), road elements (Lisini *et al.*, 2006) and/or road networks (Bentabet *et al.*, 2003), and so on. Moreover, although a number of approaches meant for SAR image analysis has been available in technical literature, no approach is likely to introduce all the spatial and spectral features that are needed for a cartographic feature extraction process starting from SAR data. For instance, road extraction can be found in many works, but this is seldom connected to urban area extraction and the use of different strategies according to the urban or non-urban areas (see Tupin *et al.*, 2002, or Wessel, 2004). The same is true for the reverse approach.

It is therefore interesting to try and prove that an effective procedure can put together starting from (some of) these or similar algorithms, and thus exploiting as much as possible the full range of information available within a high spatial resolution SAR scene. This work is instead a first attempt to provide a comprehensive approach to SAR scene characterization paying attention to the multiple elements in the same scene. This research is based on the experience of the Remote Sensing Group of the University of Pavia in analyzing airborne SAR high spatial resolution images and on the SAR acquisition campaign over the Piemonte region, in Northern Italy, performed by Intermap Technologies using their Star-3i sensor.

The data was used for the Geographic Information System developed for the 2006 Winter Olympic Games in Turin. Additionally, a study about the feasibility of cartographic feature extraction and scene interpretation was considered as a joint effort of Intermap Technologies, the University of Pavia and the national mapping agency for Italy, the Istituto Geografico Militare Italiano (IGMI). The aim of this research was the definition of a simple yet effective approach for image analysis and cartographic feature extraction starting from SAR data.

2. THE PROPOSED PROCEDURE

A common methodology for scene interpretation is based on knowledge-based segmentation of the image into simpler elements, exploiting the relationships between objects and features. This approach, usually labelled as “top-down” analysis is also implemented in this work. The novelty of the analysis proposed in this work is the contemporaneous exploitation of spectral and spatial features. Spatial features are here referring to both texture analysis and linear element extraction and recombination, which allows a better characterization of the elements in the scene than each of the two spatial analysis taken alone. Moreover, specific approaches are introduced for different parts of the scene, and associated spatial features are chosen accordingly.

The overall structure of the procedure is presented in fig.1, where solid lines represent computational steps, while dashed lines relationships. As said, this procedure exploits numerous spatial and spectral features, namely

- spectral and spatial features to extract river/water bodies,
- linear features to be grouped into the main road network,
- texture features for delineation of human settlements and urbanized areas,
- linear feature extraction, junction characterization and urban road network delineation,
- texture features for discrimination of vegetated areas along water bodies,
- statistical features for extra-urban areas segmentation and analysis of different cultivations.

The procedure has been devised for SAR images. When InSAR information is available, two additional steps are implemented:

- discrimination of low-rise and high-rise building

inside the boundary of human settlements, using 3D features;

- refinement of urban roads by data fusion of two-dimensional and three-dimensional data, using a combination of linear and 3D features.

More precisely, some of the procedures implemented for extracting the above mentioned elements are summarized in the following paragraphs.

- **Water bodies extraction:** water bodies are characterized by homogeneous or low textured areas, with low backscattering values. Moreover, their shape is smooth and regular. Therefore, “low” backscattered values are considered, and a shape regularization algorithm based on Gamba *et al.* (2007) is implemented by a reduction of the “irregularities” of the borders due to misclassifications and considering spatial relationships with other classes in the same environment.
- **Human settlement delineation:** heavily textured areas are connected to human settlement usually, but relationships with water bodies are to be considered, which allows for instance discarding highly textured area long the rivers because they most probably are woods or isolated groups of trees. Furthermore, constraints on the kernel for texture extraction can be considered, to take into account the scale of these settlements.
- **Extraction of the main road network:** roads extraction in the suburban context can be computed using a spatially reduced version of the image. A scale reduction of about 1/3 is an indicative value. In high-resolution SAR images roads are no more a subset of image edges. Instead, they usually appear as dark elongated areas with bright lateral edges. Therefore, one may detect roads by looking for pairs of parallel

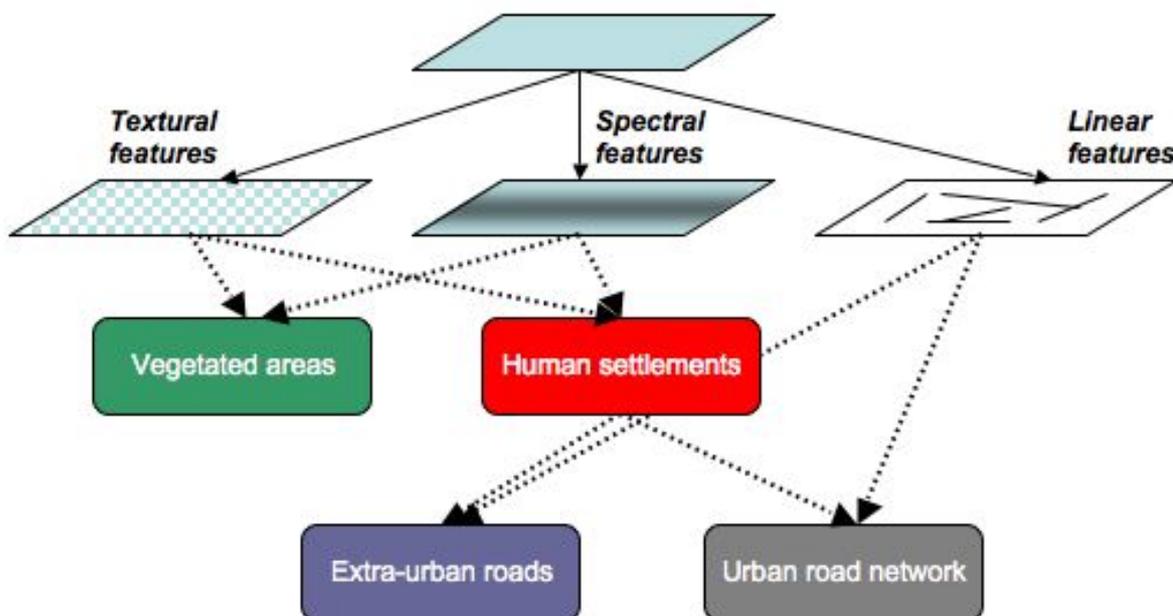


Figure 1. Overall structure of the proposed procedure.

edges or searching for dark homogeneous areas. Both of these methods, however, are subject to false positives (e.g., other artificial structures and low-reflectance areas, respectively). A more precise approach may be one using a combination of these ideas. This is the aim of the algorithm used in this context (Dell'Acqua *et al.*, 2003), which furthermore integrates road features into a multiscale-feature fusion framework whose results will be further elaborated by an alignment routine (Dell'Acqua *et al.*, 2005).

- **Extraction of urban road network:** inside the human settlement areas delineated in one of the previous step, more precise linear feature can be used to extract the road network. Here not only road candidate, but also junctions are considered according to (Negri *et al.*, 2006). Scale factors are also different than for extra-urban analysis and the full resolution is to be considered.
- **Extraction urban roads through DSM/DTM:** in case of availability of SAR and InSAR data, which is the case for the available data set, it is possible to improve the extraction of roads in urban area. Indeed, a computation of DTM allows finding areas raising above the ground. Ground areas are parks and roads. While the first ones are easily discriminated due to their shapes, candidate roads may be injected in the above found urban road network to improve the overall results.

3. EXPERIMENTAL RESULTS

As stated in the introduction, the proposed procedure was applied to a portion of the whole Piemonte data sets, recorded by Intermap Technologies and released to IGMI. The available portion consists in one IGMI tile, i.e. a 30 km by 40 km area covering a portion of the Southern part of the Piemonte region. An IGMI tile correspond to more (exactly, 16) Intermap tiles, each one covering a 4800 by 6000 pixel area, corresponding to 45 square km. The SAR data were provided in georeferenced format, with spatial posting of 1.25 m. Coregistered to the intensity two-dimensional information, three-dimensional interferometric data were also provided. Together with the raw 3D data, labelled as Digital Surface Model (DSM), a Digital



Terrain Model (DTM), obtained by Intermap technologies by means of proprietary software and procedures was also provided, featuring the terrain height, as well as corrected 3D information for water bodies. DTM and DSM have twice the spatial resolution than intensity data, and are thus provided on a 2.5 m spaced grid.

To provide a working example of the procedure proposed in section 2, in the following the intermediate elaboration results for one sample of an Intermap tiles are proposed and results are discussed. This would be useful, for instance, to highlights the advantages but also the problems of the proposed semi-automatic feature extraction and combination approach. In turn, this could also be useful for detecting new research or implementation lines for the future.

Fig. 2(a) shows the original 2D data, while fig. 2(b) provides a bird's eye of the area, as obtained overlaying the SAR intensity information of the corresponding DSM. The area covers a portion of the path of the river Po, the major river in Northern Italy, and the mostly rural area surrounding the river. In particular, the small settlement in the areas is named Ceretto.

As discussed in previous section, the first step of the procedure is the extraction of any existing water body in the scene, based simply on backscatter analysis. Since thresholding is not going to be effective due to speckle noise, the following shape analysis discards false positives and redefines the overall shape of the features by smoothing erroneous detection results on the boundary area between water and soil. Fig. 3(a) and (b) refer to these two intermediate results. According to the procedure, next step is human settlement extent delineation. Technical literature agrees that the best approach for this task is based on texture-based discrimination. The problem is the choice of the texture and the corresponding scale, which depends on the spatial resolution of the data and the settlement spatial structure. Multi-scale analysis would be more effective, as well as multiple direction for oriented textures. A suitable combination of these features is indeed able to obtain "hints" for human settlements (Pesaresi *et al.*, 2007), to be further specialized to find urban extents. In this work, we apply morphological closing to fill in the gap and obtain the boundaries of areas that are mostly likely to be human settlement or man-made elements of the landscape (with the exception of roads and other transportation infrastructure, which are oriented and can be discarded using this assumption). Fig. 3(c) shows a simple anisotropic texture information, the data range, which highlights human settlement hints, while fig. 3(d) provides a mask delineating what are likely to be human settlements.

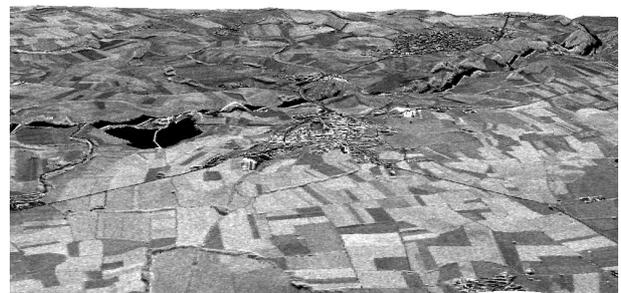


Figure 2. Sample HR SAR amplitude image and corresponding interferometric DSM in 3D view.

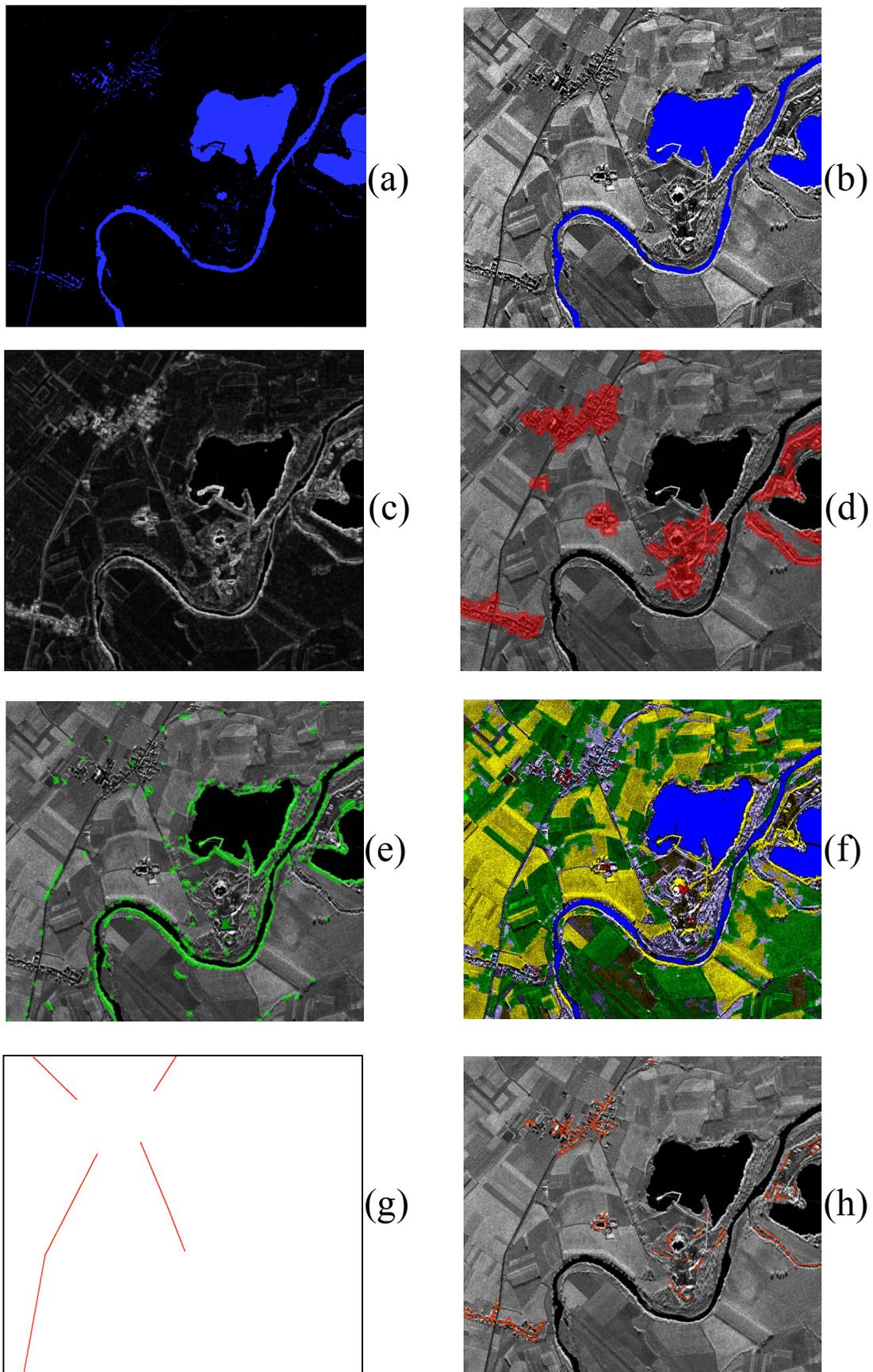


Figure 3. (a) water body extraction by amplitude SAR data thresholding; (b) refined water bodies using shape information to improve previous extraction results; (c) data range image; (d) human settlements extracted from previous image; (e) vegetation; (f) crops (two classes); (g) main road outside the human settlements; (h) road network inside the settlements.

As noted above, roads and other transportation infrastructures, although man-made landscape features, can be easily discriminated because of their geometrical characteristics, which could be coupled in HR SAR images with typical radar responses as in Negri and Gamba (2006). Moreover, they tend to be connected in a topologically consistent network, and this could be used to improve the first step, i.e. element extraction, by inserting missed elements and discarding false positives. In order to retrieve main road elements, a suitably downsampled image is considered, and the procedure proposed in Dell'Acqua *et al.* (2005) is applied. Results are shown in fig. 3(g) while in fig. 3 (h) the same procedure, but with full data resolution and stronger constraints on road network connectivity has been applied to the portion of image labelled as "human settlement". Road density might be used, in the end, as a further validation of the human settlement hypothesis or to refine the urban extent boundary. Finally, vegetation mapping is considered. Woods are extracted by the assumption that they are characterized by means of a different textural feature set than human settlement, due to the scale and physical nature of the elements in the texture (trees instead of buildings). The procedure exploits therefore the same data range information, but with different scale and threshold values. Moreover, proximity to the water bodies is considered as another hints for woods and trees. As for other vegetation types, SAR amplitude data may be able to extract boundaries between crops, if a suitable segmentation algorithm is applied. After a despeckling procedure, therefore, a Markov Random Field approach, well validated by most recent technical literature (Huawu and Clausi, 2005) is used, and corresponding classes of crops are extracted.

At the end of this procedure, there a clear need for evaluating its results. However, the only independent validation of these results is obtained by comparison with the existing Regional Technical Maps and the corresponding GIS layers of the area provided by the competent regional administration. As shown in fig.4, the Regional Technical Map and these layers are far more detailed than the extracted maps. It is however worth noting that, qualitatively speaking, the results of the semi-automatic procedure are good. Indeed, the main features of the scene have been correctly extracted, the change of the river path correctly detected and the main road network is where it is expected. A quantitative evaluation of the results in fig. 4(a) is underway. The cartographic features extracted in this as well as another sample area are going to be validated by means of a ground campaign by IGMI later this year.

As a final example, fig. 5 shows the scene results for two complete Intermap tiles. In particular, the map on the right refers to an urban-rural fringe (referring to the towns of Nichelino and Candiolo) has been enriched by detecting the road overpasses (red dots). As stated above, this is an example of high-level cartographic features easy to obtain by combining two- and three-dimensional features. In this case, road junctions, detected starting from extracted road elements, are combined with terrain height to validate or discard the "overpass" hints based on junction detection.

4. CONCLUSIONS

This work shows that a suitable combination of feature extraction algorithms, specifically developed for high spatial resolution SAR data, can be combined to obtain an effective mapping chain. It includes both two-dimensional and three-dimensional feature extraction, and possibly 2D and 3D feature

combination. The main outcomes of this work are therefore mainly the following two points.

- 1) There exist a few specific processing tools developed for SAR scene interpretation, able to jointly consider spectral and spatial, as well as context information. These tools provide effective mapping results for HR SAR in both rural and urban areas.
- 2) The exploitation of HR SAR for mapping purposes can be based on the competitive advantage that interferometric SAR provides at once both the two-dimensional and the three-dimensional representation of the same scene. A combined use of both data sets, by their same nature already co-registered, is able to improve purely 2D mapping results. It can also detect features that are more complex to detect from aerial optical data just because of the lack of 3D information.

REFERENCES

- Askne, J.; Santoro, M.; Smith, G.; Fransson, J.E.S.; Multitemporal repeat-pass SAR interferometry of boreal forests. *IEEE Trans. Geosci. Remote Sens.*, 41(7), pp. 1540-1550.
- Bentabet, L.; Jodouin, S.; Ziou, D.; Vaillancourt, J, 2003. Road vectors update using SAR imagery: a snake-based method. *IEEE Trans. Geosci. Remote Sens.*, 41(8), pp. 1785-1803.
- Dell'Acqua, F., P. Gamba, G. Lisini, 2003. Road map extraction by multiple detectors in fine spatial resolution SAR data. *Canadian Journal of Remote Sensing*, 29(4), pp. 481-490.
- Dell'Acqua, F., P. Gamba, G. Lisini, 2005. Road extraction aided by adaptive directional filtering and template matching. In: Proc. of *URBAN 2005*, Tempe AZ, USA 14-16 March 2005, unformatted CD-ROM.
- EUSAR06. www.dlr.de/hr/eusar2006 (accessed 10 June 2007)
- Gamba, P., F. Dell'Acqua, G. Lisini, 2007. Raster to Vector in 2D Urban Data. In: Proc. of *JURSE 2007*, Paris, France, 13-15 April 2007, unformatted CD-ROM.
- Hall, O.; Falorni, G.; Bras, R.L., 2005. Characterization and quantification of data voids in the shuttle Radar topography mission data, *IEEE Geoscience and Remote Sensing Letters*, 2(2), pp. 177-181.
- He, C.; Xia, G.-S.; Sun, H., 2006. An adaptive and iterative method of urban area extraction from SAR images. *IEEE Geoscience and Remote Sensing Letters*, 3(4), pp. 504-507.
- Huawu D.; Clausi, D.A., 2005. Unsupervised segmentation of synthetic aperture Radar sea ice imagery using a novel Markov random field model, *IEEE Trans. Geosci. Remote Sens.*, 43(3), pp. 528-538.
- IGARSS'06. <http://www.igarss06.org> (accessed 10 June 2007)
- JURSE2007. <http://tlc.unipv.it/urban-remote-sensing-2007/> (accessed 10 June 2007)
- Lisini, G.; Tison, C.; Tupin, F.; Gamba, P., 2006. Feature fusion to improve road network extraction in high-resolution SAR images. *IEEE Geoscience and Remote Sensing Letters*, 3(2), pp. 217-221.

OBJECT EXTRACTION AT AIRPORT SITES USING DTMs/DSMs AND MULTISPECTRAL IMAGE ANALYSIS

N. Demir*, E. Baltsavias

ETH Zürich

Institute of Geodesy and Photogrammetry,
Wolfgang Pauli Str. 15, CH-8093 Zurich, Switzerland

E-mail: (nusret.demir, manos)@geod.baug.ethz.ch

KEYWORDS: DTMs/DSMs, LIDAR Data Processing, Multispectral Classification, Image Matching, Information Fusion, Object Extraction

ABSTRACT

The automatic detection and 3D modeling of objects at airports is an important issue for the EU FP6 project PEGASE. PEGASE is a feasibility study of a new 24/7 navigation system, which could either replace or complement existing systems and would allow a three dimensional truly autonomous aircraft landing and take-off primarily for airplanes and secondary for helicopters. This new navigation system relies on three key technologies:

- The specification and acquisition of a reliable geospatial reference database of the airports.
- Various sensors onboard the aircraft that perform a real-time extraction of relevant features from the data.
- Innovative correlation techniques between the above features and the onboard reference database to determine the location of the aircraft and plan its future trajectory for safe landing or take-off.

In this work, we focus on the first topic. Since often existing data for airports have not the necessary accuracy, resolution and/or currency, we need to develop automated methods for generating them.

1. INTRODUCTION

The automatic detection and 3D modeling of objects at airports is an important issue for the EU FP6 project PEGASE.

PEGASE is a feasibility study of a new 24/7 navigation system, which could either replace or complement existing systems and would allow a three dimensional truly autonomous aircraft landing and take-off primarily for airplanes and secondary for helicopters. This new navigation system relies on three key technologies:

- The specification and acquisition of a reliable geospatial reference database of the airports.
- Various sensors onboard the aircraft that perform a real-time extraction of relevant features from the data.
- Innovative correlation techniques between the above features and the onboard reference database to determine the location of the aircraft and plan its future trajectory for safe landing or take-off.

In this work, we focus on the first topic. Since often existing data for airports have not the necessary accuracy, resolution and/or currency, we need to develop automated methods for generating them. As a first object of focus, we deal with vegetation (mainly trees), and to a lesser extent buildings and other objects.

2. INPUT DATA

We have worked on the area of Zurich airport. The available data for this region are:

- Vector map
- Color and CIR image data
- Orthophotos
- LIDAR DSM/DTM raw point cloud
- LIDAR DSM/DTM GRID surface model data

2.1 Vector data:

The vector map has 20 cm horizontal accuracy. It has been produced using semi-automated measurement of buildings and other objects (poles, towers, etc.) from stereo aerial images using CC-Modeler.

2.2 Image data:

Color and CIR (color infrared) image data have been used for generation of DSM and multispectral classification process. Metadata of images can be seen in Table 1.

Film	Color and IR
Camera Type	Frame
Focal Length	303.811 mm.
Scale	1:10'150 (color) / 1: 6'000 (IR)
Endlap	70%
Sidelap	26%
Scan resolution	20 micron

Table 1. Image data description

Orthophoto of the airport region has been produced using an image matching DSM (see below) with 1m resolution.

2.3 LIDAR Data:

LIDAR DSM-AV and DSM point cloud data have been used for extraction of objects process. The process can be seen at the section 4. The DSM-AV point cloud includes all lidar points with 4 echoes per pulse. The DTM-AV data includes only points on the ground, so it has holes at building positions, and less density at tree positions. The altimetric precision (one standard deviation) of both models is 0.5 m, and 1.5 m at vegetation and buildings. The density is approximately 1 point by 2 m². The GRID DSM and DTM have a grid spacing of 2m

and same accuracy specifications as the point cloud models. They represent the visible and terrain surface respectively.

3. OBJECT EXTRACTION FROM DIGITAL SURFACE MODELS

3.1 Extraction of buildings using SCOP++

SCOP++ software derives from a point cloud a DTM and DSM and classifies the original points using a hierarchical method. In each hierarchy level, robust interpolation for the classification of the points and the surface derivation is done. SCOP++ performs filtering of airborne laser scanning data for automatic classification of the raw point cloud into terrain and off-terrain points, i.e. for extracting the true ground points for further DTM processing. It uses efficient robust interpolation techniques with flexible adaptation to terrain type and terrain coverage. The package has user-controlled hierarchical process using point cloud pyramids and elimination of gross errors (Kraus et al., 1998, Pfeifer et al., 1998).

We applied the automatic LIDAR classification tool and it performed the separation process as building, vegetation and ground points for the DSM raw lidar point cloud (Fig. 1).

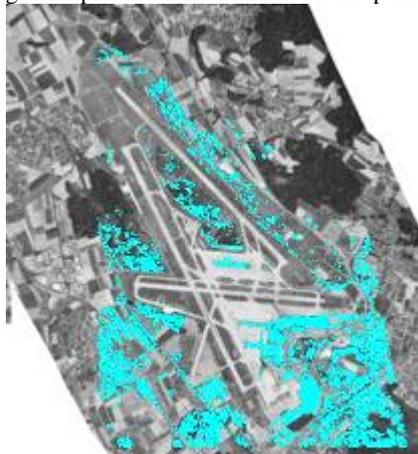


Figure 1. First results for buildings on reference orthophoto

For refining of results, the DTMaster programme has been used. This software gives the opportunity to improve the result quality using orthophotos overlapping with the point classes. Manually selected points can be easily moved to another class (vegetation, ground) or be deleted interactively. Final building points have been refined and the result can be seen in Fig. 2.

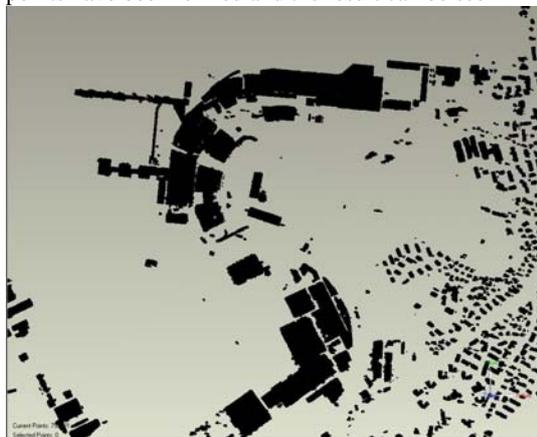


Figure 2. Results from SCOP++/DTMaster

The result of classification process is partly unknown and we tried to develop additional procedures for extraction of buildings and trees.

3.2 DSM generation using image matching

At the Institute of Geodesy and Photogrammetry, ETH Zurich, a full suite of new algorithms and the software package SAT-PP (Satellite Image Precision Processing) have been developed for the precision processing of high-resolution satellite and aerial image data. The software features include: GCP measurements, image georeferencing with RPC approach and other sensor models, DSM generation with advanced matching techniques, orthoimage generation, semi-automated feature extraction etc. (Gruen et al., 2005). The mathematical description of the sensor model and block adjustment developed in SAT-PP is given in Baltasvias et al. (2005) and Gruen et al. (2005).

11 aerial images of the area have been used as input. The matching approach used a coarse to fine hierarchical solution with a combination of multi image matching algorithms and automatic quality control.

In order to reduce the effects of the radiometric problems such as strong bright and dark regions and to optimize the images for subsequent feature extraction and image matching, a preprocessing method, which combines an adaptive smoothing filter and the Wallis filter, was developed. Firstly, an adaptive smoothing filter is applied to reduce the noise level and to sharpen edges and preserve even fine details such as corners and line end-points. Next, the Wallis filter is applied to enhance and sharpen the already existing texture patterns. After the image preprocessing, the image pyramid is generated starting from the original resolution images. Three additional pyramid levels are generated, each with a reduced resolution by factor 3 from the previous level.

After pre-processing of the original images and production of image pyramids, the matches of three kinds of features (feature points, grid points and edges) in the original resolution images were found, progressively starting from low-density features on the lower resolution levels of the image pyramid. Seed points were selected by stereo-measurement technique. Finally, we generated an interpolated regular DSM with 2 m resolution for the airport area (Fig. 3).



Figure 3. Matching DSM for Zurich Airport

3.3 Extraction of vegetation areas using multispectral classification and DSM

A CIR orthophoto was produced using the matching DSM for high vegetation extraction (Fig. 4). When calculating a DSM from aerial images, trees produce the same effect as buildings, showing up as features above ground (Niederöst, 2001). Applying classification process using the orthophoto gives the

tree region. We applied this process on a small part of the airport site.

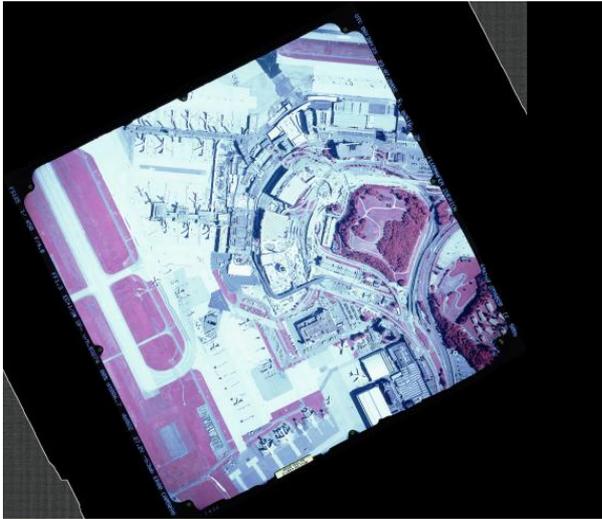


Figure 4. Orthophoto

We used supervised classification which uses NDVI analysis as a traditional vegetation method. The spectral characteristics of vegetation can help to separate it from other objects. Near infrared(NIR), red(R) and green(G) channels were used to classify vegetation areas using the normalised vegetation index given as $NDVI = (IR-R)/(IR+R)$. It takes values between [-1;1]. Our main target classes for airports are open grass, high vegetation (trees), man-made objects (mainly buildings, roads, poles). Here, we focused to extract mainly trees using multispectral information.

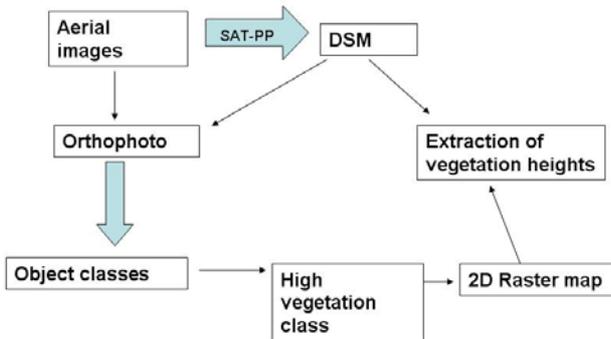


Figure 5. Workflow of vegetation extraction

The training areas were selected manually using AOI tools with the ERDAS Imagine commercial software. Automatic supervised classification produced 3 classes: open grass, trees and man-made structures (see Fig. 6).

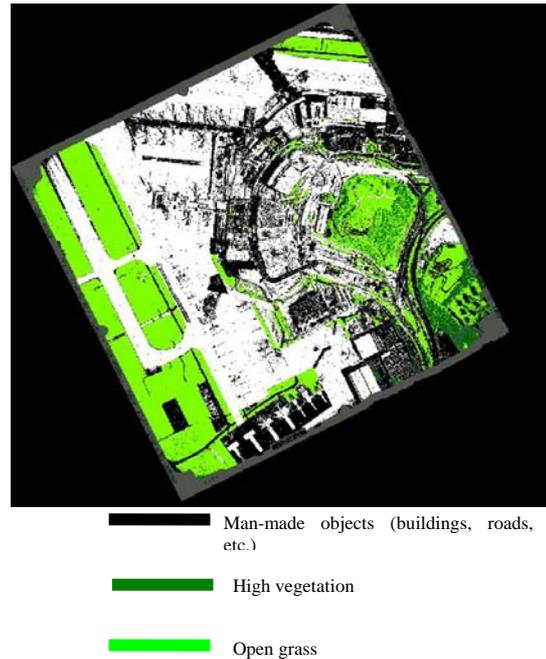


Figure 6. Classification result for vegetation extraction.

The region which was classified as high vegetation, was converted to a 2D raster map with 1m resolution, using pattern recognition and image processing techniques. Tree heights were interpolated in the DSM (Fig. 7). The workflow of the extraction process can be seen in Figure 5.



Figure 7. Extracted vegetation and digital surface model

3.4 Extraction of objects using existing vector maps and DSM

In the past, many researchers studied the extraction of objects using existing plans. Haala et al. (1998), Brenner (2000), Vosselman and Dijkman (2001), Hofmann et al. (2002), Hofmann (2004), Vosselman (2003), Koch (2004), Overby et al. (2004), Stoter and Oosterom (2004) used 2D maps as prior information for extracting buildings by using LIDAR data. Haala et al. (1998) described a method that combines height data provided by airborne laser scanning and existing ground plans of buildings in order to enable an automatic data capture by the integration of these different types of information. Vosselman and Dijkman (2001) focused on the extraction of the roof faces and the generation of 3D building models by

combining the extracted roof faces with the ground plans. Hofmann et al. (2002) transformed laser data into raster format, in order to extract buildings from laser data. Vosselman (2003) described several algorithms and procedures developed for the 3D reconstruction of streets and trees from airborne laser altimetry data in combination with a cadastral map. Overby et al. (2004) generated roof polygons using airborne laser scanning and ground plans (footprints) extracted from technical feature maps and applied plane extraction by a 3D Hough transform.

We used the existing vector map and converted the building polygons to a 2D raster map (Fig.8).

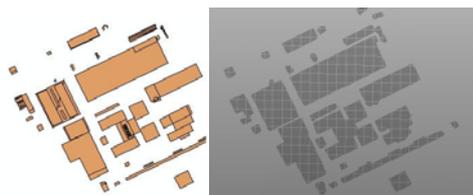


Figure 8. Object polygons to 2D raster map

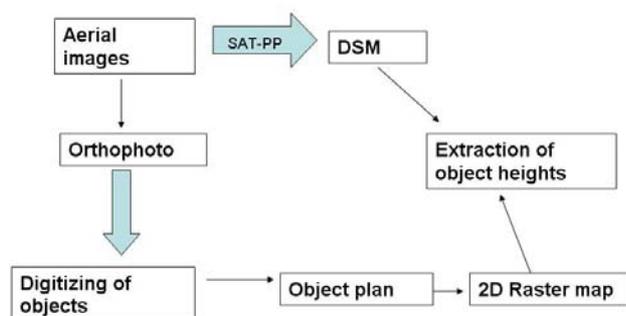


Figure 9. Extraction of objects using object vector maps

The height for the 2D building points (see Fig.9) was interpolated from the DSM (Fig.10).



Figure 10. Extraction of buildings and DSM

4. EXTRACTION OF OBJECTS USING DENSITY CALCULATION OF LIDAR DATA

When no prior existing data are available, exploring other methods for detection of objects can be useful. One of these methods is using the density information of LIDAR based DTM and DSM. As input data, we use the raw lidar data and

multispectral aerial images from the Zurich airport. They include 2 datasets acquired with leaves off: a) terrain-only points (DTM-AV) (see Fig. 11), b) all points with 4 echoes per pulse (DSM-AV). Thus, the last dataset shows the structure of trees (see Fig. 12).



Figure 11. DTM-AV Data for the Zurich Airport, holes in the data points mainly at buildings

Using high-resolution LIDAR data, it is not only possible to detect buildings and their approximate outlines, but also to extract planar roof faces and, thus, to create models of the roof structure. With decreasing resolution, it becomes more difficult to correctly detect buildings in LIDAR data, especially in residential areas characterised by detached houses (Rottensteiner et al., 2003).



Figure 12. Profile tree view from LIDAR DSM raw data

We tried to develop an own method for detecting buildings, using the multiple echoes for each pulse, and the 3D nature and higher point density of lidar point clouds with trees. Buildings can be detected from the DTM-AV raw dataset using the fact that there are no measurements at their planimetric position. Indication of trees can be used, by using the image classification results as a mask and using the raw DTM-AV data. Using the same mask, with the raw DSM-AV data, will show that at tree areas, the point density is much higher than at open terrain or buildings, so this indicator could be used.

4.1 Calculation of point density

This algorithm calculates total number of points in $2m^2$ area squares of the DTM-AV data. The number of points in per $2m^2$ was reported to a file (Table 2).

This file contains three columns. Two columns show the searched position and one column presents the total number of points.

Number of points	X	Y
10	679332	234366
6	679332	234368
2	679332	234370
0	679332	234372
.	.	.
.	.	.

Table 2. Density report

X and Y columns present the 2m range scale. For example if there is a value as 679332, it presents a 2m range between 679330-679332 values.

4.2 Extraction process

Another algorithm searches the no-point areas using the above density report file and extracts these areas from the DSM. Firstly, the code searches the columns of “number of points” and if the value is 0, then it extracts the corresponding range area from digital surface model.

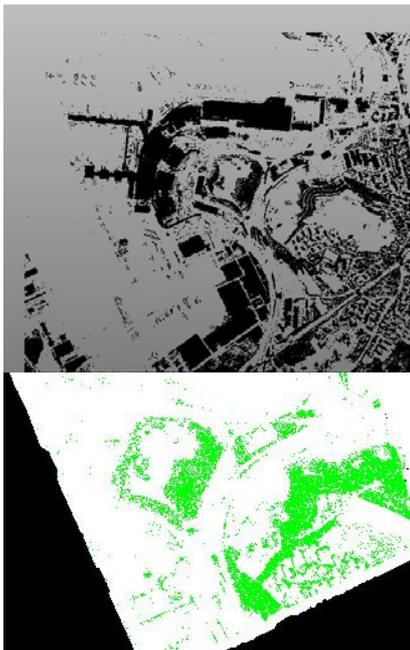


Figure 13. Extracted objects (top) and vegetation mask (bottom)

First results show mainly buildings and secondly high trees and some objects (airplanes, cars, etc.). Using the classification mask with this the result can provide the elimination of tree points from other points (Fig. 13).

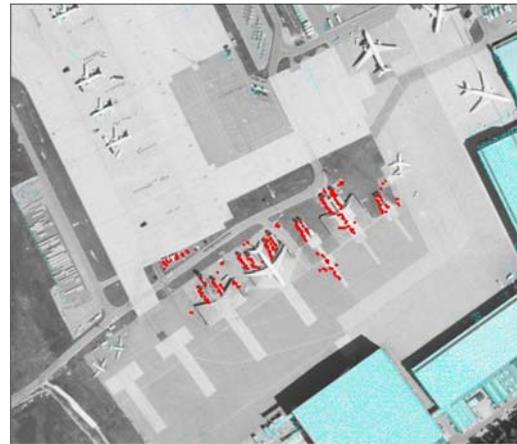


Figure 14. Selected other-than-building object points (e.g. airplanes)

The final result was improved using DTMaster software with manual elimination of tree points, airplanes etc. (Fig. 14-15).

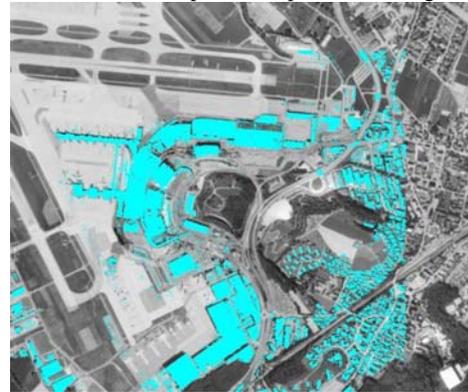


Figure 15. Final extracted building points

4.3 Performance evaluation

We compared the results from SCOP++ processing and density calculation process (Figures 16 and 17). Because of the low density of the DTM-AV data, small buildings were not detected as well as with the SCOP++/DTM edit package classification. Large airport buildings could be extracted equally well as with the other methods.

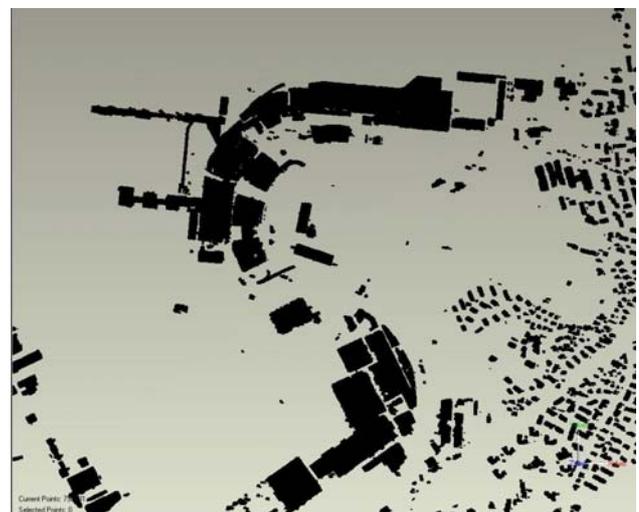


Figure 16. Results from SCOP++/DTMaster

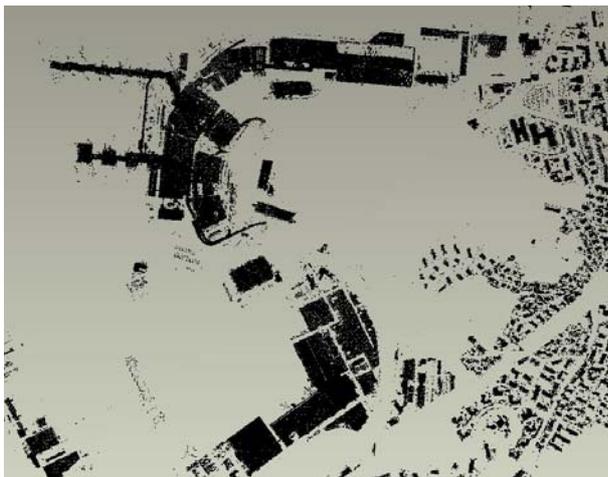


Figure 17. Results from density calculation

If we compare the results from the rasterised vector maps and the SCOP LIDAR/ DTM edit package, we can see that the results are very similar (Fig. 18).

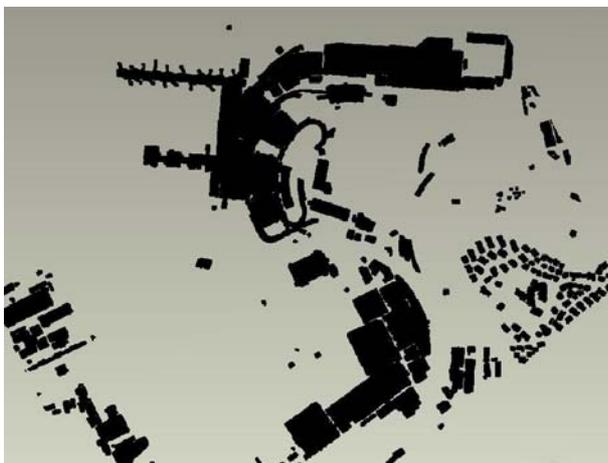


Figure 18. Results from the rasterised vector maps

5. CONCLUSIONS

We have presented four methods for object extraction for airport sites. These methods are using commercial LIDAR processing software, multispectral classification using orthophotos, calculation of point density of DTM-AV and rasterised vector maps. When LIDAR DTM-AV and DSM data are available and if commercial software like SCOP is unavailable, point density of raw DTM and DSM data can be used mainly for large object extraction. When LIDAR data is unavailable, DSMs can be generated using multi-image matching methods (e.g. SAT-PP software). The use of multiple cues, especially multispectral information, point density, DSM, 2D location of reference data support the building extraction result. Future work will include the combination of these various cues, and quantitative evaluation using the existing reference data.

6. REFERENCES

Baltsavias, E., Zhang, L., Eisenbeiss, H., 2006. DSM Generation and Interior Orientation Determination of IKONOS Images Using a Testfield in Switzerland. *Photogrammetrie, Fernerkundung, Geoinformation*, (1), pp. 41-54.

Brenner, C., 2000. Towards fully automatic generation of city models. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 33, Part B3/1, pp. 85-92.

Gruen, A., Zhang, L., Eisenbeiss, H., 2005. 3D Precision Processing Of High Resolution Satellite Imagery, *Proceedings of ASPRS 2005 Annual Conference*. (only on CD-ROM).

Haala, N., C. Brenner, Anders, H., 1998. 3D Urban GIS From laser altimeter and 2D map data. *Journal of Photogrammetry & Remote Sensing*, Vol. 32, Part 3/1, pp.339-346

Hofmann, A., Maas, H.-G., Streilein, A., 2002. Knowledge-Based building detection based on laser scanner data and topographic map information. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol 34, Part 3A+B pp. 169-175.

Hofmann, A. D., 2004. Analysis of TIN-structure parameter spaces in airborne laser scanner data for 3-d building model generation. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol.35, Part B3, pp.302-307.

Koch, A., 2004. An approach for the semantically correct integration of a DTM and 2D GIS vector data, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* Vol. 34, Part B4, pp.523-529.

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, pp. 193-203.

Niederöst, M., 2001. Automated update of building information in maps using medium scale imagery, *Automatic Extraction of Man-Made Objects from Aerial and Space Images (III)*, E. Baltsavias, A. Gruen, L. Van Gool (Eds.), Balkema, Lisse, pp. 161-170.

Overby, J., Bodum, L., Kjems, E., Iisoe, P.M., 2004. Automatic 3D building reconstruction from airborne laser scanning and cadastral data using hough transform *International archives of photogrammetry and remote sensing*, Vol. 35, Part. B3, pp. 296-301.

Pfeifer N.; Kostli A., Kraus K., 1998. Interpolation and filtering of laser scanner data – implementation and first results. *International archives of photogrammetry and remote sensing*, Vol 37, Part 3/1, pp. 153 - 159.

Rottensteiner, F., Trinder, J., Clode, S., Kubik, K., 2003. Detecting buildings and roof segments by combining LIDAR Data and multispectral images. In *Proceedings of Image and Vision Computing Symposium New Zealand (IVCNZ 2003)*, pp. 60-65.

Stoter, J., Oosterom, P., 2005. Technological aspects of a full 3D cadastral registration. *International Journal of Geographical Information Science* Vol. 19, pp. 669-696.

Vosselman, G., Dijkman, S., 2001. 3D Building Model reconstruction from point clouds and ground plans. *International archives of photogrammetry and remote sensing*, Vol. 34, Part 3, pp. 37-43.

Vosselman, G., 2003. 3D Reconstruction of roads and trees for city modeling. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 34, Part 3/W13, pp. 231-236.

Zhang, L., Kocaman, S., Akca, D., Kornus, W., and Baltsavias, E., 2006. Test and performance evaluation of DMC images and new methods for their processing. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* Vol. 36, Part 1 (only on CD-ROM).

AN APPROACH FOR SEMI-AUTOMATIC EXTRACTION OF FEATURES FROM AERIAL PHOTOGRAPHS

O. Eker¹, A.C.KIRACI², D. Z. Şeker³

¹Harita Genel Komutanlığı, Fotogrametri Dairesi Başkanlığı, Ankara, oktay.eker@hgk.mil.tr

²Harita Genel Komutanlığı, Fotogrametri Dairesi Başkanlığı, Ankara, alicoskun.kiraci@hgk.mil.tr

³İstanbul Teknik Üniversitesi, Jeodezi ve Fotogrametri Mühendisliği Bölümü, Fotogrametri Anabilim Dalı, İstanbul, dzseker@ins.itu.edu.tr

KEYWORDS: Photogrammetry, automatic feature extraction, semi-automatic digitizing, image segmentation, level set algorithms, conversion from raster to vector.

ABSTRACT:

Aerial photographs have been evaluated manually by the operators for the extraction of the vector data to produce photogrammetric maps. In the recent years the developments, in the photogrammetry, provides to perform these extraction processes automatically.

In this study, a new semi-automatic feature extraction approach, based on the segmentation of the images using color-differences of the pixels and the propagation of fronts by the Level Set algorithms, is developed. An object-oriented application software is also developed to test the capabilities of the developed method. Some semi-automatic feature extraction applications are made by the help of the developed software using 1:4000 and 1:35000 scale black/white aerial photographs for determining the capabilities of this method.

The results of the tests show that this method can be used for the extraction of some features from aerial photographs for GIS and the production of the photogrammetric maps.

1. INTRODUCTION

Vector data which is needed to produce photogrammetric map is being digitized manually by operators from aerial photographs. These processes can be done automatically by the help of the developments in photogrammetry in recent years.

One of the methods for automatic and semi-automatic feature extraction and classification is image segmentation. Image segmentation is often used method on feature extraction and interpretation from medical images. (Adalsteinsson and Sethian, 1995). Level Set and Fast Marching methods are successfully used in recent years to provide surfaces to develop and propagate on the image taken by classification (Malladi vd., 1994).

In this paper; a new approach based on using image segmentation and level set algorithms together to extract some features semi-automatically from digital aerial photographs and software for testing the approach is introduced.

2. SEMI – AUTOMATIC FEATURE EXTRACTION APPROACH AND DEVELOPED SOFTWARE

The approach that has been developed for semi-automatic feature extraction has been based on level set and image segmentation algorithms. According to this method, three problems had to be solved. The first one was how the algorithm would be started. This problem was solved by marking any point (pixel) on the feature wanted to be digitized by the operator. Thus, level set algorithm would start to work from the point that operator chooses. This solution brought out the approach to be semi-automatic.

Second problem was which criteria would be taken in whether the digitizing feature process would progress or not. This problem was solved by making use of color values of each pixel. The color value of marked point or determined neighborhood level, color values of neighboring pixels and computed color value by getting average, is compared with color value of neighboring pixels and if the color difference is within the tolerance value determined before, algorithm is continued otherwise it is stopped. In other words, segmentation is carried out according to color differences.

After completing the determination of the feature wanted to be digitized and marking process it is needed to get the feature as vector data. Thus the data can be used in a Geographic Information System (GIS). This was the third problem. Raster to vector transformation was carried out to solve the problem. Semi-automatic digitizing approach that solves those three problems mentioned above was configured in five main steps as the integrity of processes. This process steps are:

- The selection of starting point or pixel of the feature wanted to be digitized by the operator.
- Performing the image segmentation of the selected pixel by making use of color difference with neighboring pixels.
- Propagation of the segmentation by using the level set method and storing as a heap sort.
- Getting 1 bit (1 color) masked image from those structured pixels.
- Getting features' vector data in a known format after raster to vector conversion from masked image.

First four process steps were compiled using Borland C++ programming language and using no library files except for system libraries. Fifth step was carried out by the help of a raster to vector conversion open code that is shared in the

internet. This code was developed and compiled by using Visual C++ programming language. Some functions like weed tolerance and coordinated transformation tools were added to this code.

2.1 Image Segmentation Algorithm

Image segmentation has been based on color difference and this algorithm had been made flexible by adding a tolerance value that could be set by operator. Thus operator can detect and digitize large areas at one click of the mouse by giving high tolerance value in high contrasted images. Segmentation algorithm threshold value was determined as two times of multiplied value of image width and length.

Segmentation is started from the selected feature point. Value of neighboring pixels and their average value are computed as the reference, according to pixel value and/or neighboring level of the feature point. If neighboring value is zero, pixel's color value is taken on the reference neighboring level is one, color value computed by averaging with the eight pixels around it is taken as the reference value (Eker 2006).

Here, color differences of the pixels in three bands (red, green, blue) are computed one by one and searched whether each is under the threshold or not.

2.2 Storing and Propagating With Level Set Algorithm

Starting value needed for propagation with level set or going along with the feature is copied with by means of the semi-automatic nature of developed method. Propagation of the surface would start from the pixel which is selected by the operator and zero level set would be defined by the location of starting pixel.

Another component that is needed for level set algorithm is the limit value providing the control of propagation. Solution for the limit was carried out by computing the color differences as mentioned above.

Latter there had been one problem to carry out that was which neighboring pixel would be chosen to continue fitting the limit value. As a matter of fact in this solution of the problem Fast Marching method may be used. The function would go on to propagate hence the pixel that has the least value (Sethian, 1998). However, what would the least value be?

In the developed method for the question of what would the least value represent to, the first pixel selected by the operator is admitted as zero level set function and each neighboring pixel (the first pixels in east, west, north, south direction) is checked by the color difference value algorithm mentioned above and the distance from the zero level set (first selected pixel) is computed for the pixels fit the condition, thus diffusing is carried on hence the pixel having the least distance.

Completing the propagation updating process should be carried out (Sethian, 1998). For updating process quadratic equations were used required for computing the difference (Sethian, 1998).

Minimum heap sort had been used in order to store and access the pixels. In the structure of minimum heap sort, the image cell having the least distance from level 0 would be at the top. Certain structure is required for preserving the heap sort. Adding the image cell to the heap, fixing the image cell in the

heap, being removed the image cell having the least distance from the heap and updating the structure on the event of every new image cell is added is required.

By means of this storing structure, effectiveness has been increased in large volume processes such as attaining to the image cells, testing and computing the propagation, marking the image cells and storing the marked image cells (Eker, 2006).

2.3 RASTER TO VECTOR CONVERSION

Raster to vector conversion is valid for 1 bit image files which include two types of data (0 or 1), (ESRI, 1997). By the help of the developed software, extracted pixels from the image are marked as colored and the others are black. This new image is saved as a masked image in bmp format.

Open Visual C++ code shared in the internet, was developed and compiled with adding additional functions to convert the acquired mask image file to vector data. The center and the border lines of the features are being converted from raster data to vector data by the help of an interactive interface by setting the connection with the main interface. A coordinated vector data is gained by entering the lower left corner coordinates and image resolution in both two dimensions (x, y) on the main interface. Besides, entering the weed tolerance with an interactive interface, it is possible to get vector data in desired smoothness (Eker, 2006).

If weed tolerance is 0 all the pixels are included in calculation without making smoothness. When the weed tolerance is increased, pixels with increasing intervals are taken into consideration instead of all pixels and vector data becomes smooth. However, it is probable the geometry of vector data may get worse if higher weed tolerance is given (Eker, 2006).

3. TEST STUDIES

Test studies were carried out with the program developed in two groups in order to determine the usefulness and effectiveness of this new approach exposed.

In the first group test study a 1:4000 scaled black/white digital

Figure 1. Applications realized with a 1:4000 scaled aerial photograph.

aerial photograph was used. Thus it was investigated whether the software would be used or not to extract some features in large scaled mapping.

In the second group study by using a 1:35.000 scaled photograph the same investigation was made for middle scaled mapping process.

3.1. Test 1

In the first test work that a 1:4000 scaled white/black aerial photos used which was scanned at 20 microns, roads and buildings were selected as the features that were as semi-automatically extracted from the image. It is shown in the Figure 1 that vector data which were obtained from the result of feature extraction studies that were done related to these selected features. Roads were vectorized from their center and edge lines. Except the obstacles as trees etc., the roads and buildings also the smallest secondary roads could be extracted clearly. Cuts which caused by the obstacles couldn't be solved

by this algorithm and these cuts have to be corrected manually by the operator.



Figure 1. Applications realized with a 1:4000 scaled aerial photograph.

3.1. Test 2

In this test work; three different features were tried to be extracted semi-automatically from a 1:35000 scaled, black/white aerial photo which was scanned at 20 microns. These features are a road with asphalt cover on surface, a wide watery stream and a footpath. Both asphalt road and watery stream were vectorized from their borders as polygons, but

footpath was vectorized from center line as a line feature (Figure 2). A few click with the mouse on the road and the stream were enough to extract these features but on the footpath it was difficult. The color difference of the footpath from the texture of the area was not too much, so the propagation was carried out by little steps with small color difference tolerances.

In this work, it is determined that the effects of the obstacles which cover the linear features (tree, car...etc.) become lesser with shrink of the little scale. Because of reduces in scale also the area which involved the pixel numbers that take part of these obstacles reduces. Thus blank quantities that cause by the obstacles take more little values.

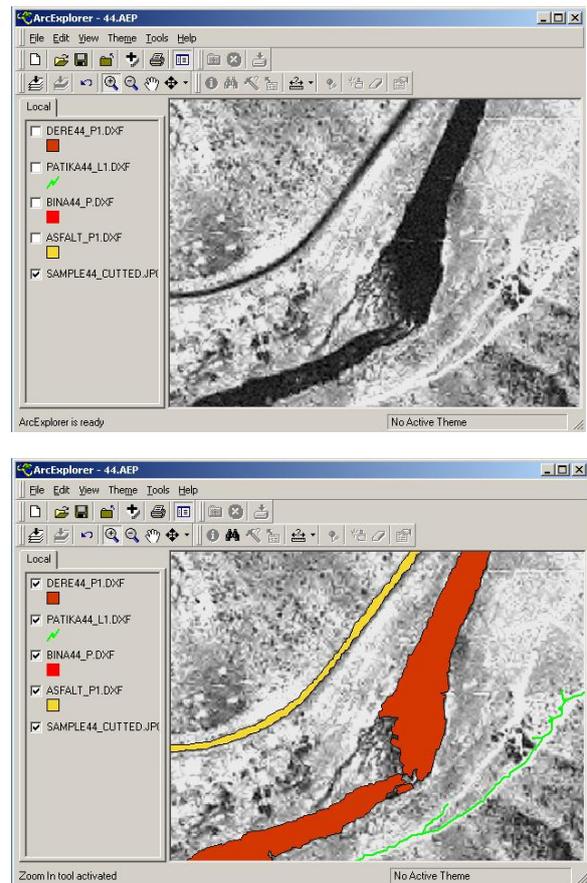


Figure 2. Applications realized with a 1:35000 scaled aerial photograph.

4. CONCLUSION

As the results; with condition of developing of the lacking aspects that are arrange in order below, this approach will be considered as useful on semi-automatic extraction of some features from image for GIS and producing photogrammetric maps from the aerial photos:

- Especially, the obstacles on the features in large scaled images negatively affect the feature extraction process.
- When the tolerance value is not properly adjusted wrong feature extractions may be occurred.
- When very large sized image files used, because of the values of pixels recorded into computer memory in this

approach, some hardware errors may be encountered because of too much memory will be needed.

- Image quality, contrast and noise ratio affect the algorithm's success in large scales.
- The linear feature's surface and cover qualities also affects the algorithm's success.

In order to be able to solve these problems and making the software more effective; for improving the contrast in images and lowering the noise ratios, filters like anisotropic diffusion filter with including ability of the edge enhancement algorithms into developed approach, for filling the blanks that are caused by the large obstacles applying ability of the different interpolation methods and for handling the big sized images in computer environment the use ability of the pyramid levels will be studied.

RESOURCES

Adalsteinsson, D., Sethian, J.A., 1995. Fast Level Set Method for Propagating Interfaces, *Journal of Comp.Phys.*, 118, 269-277.

Eker, O., 2006. . *Hava Fotoğraflarından Yarı Otomatik Olarak Çizgisel Detayların Belirlenmesi*, Doktora Tezi, İTÜ Fen Bilimleri Enstitüsü, İstanbul.

ESRI, 1997. *ARC/INFO User's Guide Cell-Based Modelling With GRID*, Redlands, USA.

Malladi, R., Sethian, J.A. ve Vemuri, B.C., 1994. Evolutionary Fronts for Topology-Independent Shape Modeling and Recovery, *Proceedings of Third European Conference on Computer Vision, Stockholm, Sweden, Lecture Notes in Computer Science*, 800, 3-13.

Sethian, J.A., 1998. Fast Marching Methods and Level Set Methods for Propagating Interfaces, *von Karman Institute Lecture Series*, Computational Fluid Mechanics, Belgium.

URL (Raster to vector conversion open code),
<http://www.xmailserver.org/davide.html> , 28 February 2006.

MULTIBASELINE INTERFEROMETRIC SAR AT MILLIMETERWAVES TEST OF AN ALGORITHM ON REAL DATA AND A SYNTHETIC SCENE

H. Essen ^a, T. Brehm ^a, S. Boehmsdorff ^b, U. Stilla ^c

^a FGAN, Research Institute for High Frequency Physics and Radar Techniques, D-53343 Wachtberg, Germany –
(essen, brehm)@fgan.de

^b Technical Center of the German Armed Forces, Oberjettenberg, D-76275 Schneizlreuth, Germany -
Stephanboehmsdorff@bwb.org

^c Technical University Munich, Arcisstr. 21, D-80290 München, Germany -
Stilla@bv.tum.de

KEY WORDS: Millimeterwaves, Interferometry, Synthetic Aperture Radar, Phase unwrapping

ABSTRACT:

Interferometric Synthetic Aperture Radar has the capability to provide the user with the 3-D-Information of land surfaces. To gather data with high height estimation accuracy it is necessary to use a wide interferometric baseline or a high radar frequency. However the problem of resolving the phase ambiguity at smaller wavelengths is more critical than at longer wavelengths, as the unambiguous height interval is inversely proportional to the radar wavelength. To solve this shortcoming a multiple baseline approach can be used with a number of neighbouring horns and an increasing baselength going from narrow to wide. The narrowest, corresponding to adjacent horns, is then assumed to be unambiguous in phase. This initial interferogram is used as a starting point for the algorithm, which in the next step unwraps the interferogram with the next wider baseline using the coarse height information to solve the phase ambiguities. This process is repeated consecutively until the interferogram with highest precision is unwrapped. On the expense of this multi-channel-approach the algorithm is simple and robust, and even the amount of processing time is reduced considerably, compared to traditional methods. The multiple baseline approach is especially adequate for millimeterwave radars as antenna horns with relatively small aperture can be used while a sufficient 3-dB beamwidth is maintained.

The paper describes the multiple baseline algorithm and shows the results of tests on a synthetic area. The test objects are representative in size for generic buildings. Possibilities and limitations of this approach are discussed. The relevance for applications of millimeterwave InSAR as means for the extraction of characteristics of buildings is highlighted. Examples of digital elevation maps derived from measured data are shown.

KURZFASSUNG:

Mit interferometrischem SAR können dreidimensionale Informationen über Landoberflächen gewonnen werden. Um Daten mit hoher Höhenschätzgenauigkeit gewinnen zu können, muss entweder eine große interferometrische Basis gewählt oder bei geringerer Basisbreite eine höhere Betriebsfrequenz verwendet werden. Bei höherer Radarfrequenz ist jedoch das Problem der Phasenfortsetzung schwieriger zu lösen, da das eindeutig bestimmbare Höhenintervall invers proportional zur Frequenz ist. Um diesen Nachteil auszugleichen, kann ein Multibasenansatz verwendet werden, der mehrere benachbarte Sende-/Empfangshörner mit zunehmender Basisbreite nutzt. Für die geringste Basisbreite sollte die Phasenbeziehung eindeutig sein. Das zugehörige Interferogramm wird als Ausgangsdatensatz für ein rekursives Verfahren genutzt, bei dem die Grobinformation des ungenaueren Interferogramms als Basis für das nächst feinere Interferogramm dient. Mit dem Nachteil des höheren Hardwareaufwandes ergibt sich hier ein sehr einfacher und dabei schnell arbeitender Ansatz für den Phasenfortsetzungsalgorithmus. Der Mehrbasenansatz ist besonders für den Millimeterwellenbereich geeignet, da hier Antennenhörner mit relativ geringer Apertur bereits eine ausreichende Bündelung bieten.

Der Beitrag beschreibt den Mehrbasenalgorithmus und zeigt Ergebnisse für eine generische Testszene. Die Bedeutung dieser Methode für die Bestimmung der Höhencharakteristik von Gebäuden wird besprochen. Vor- und Nachteile dieses Ansatzes werden diskutiert und Beispiele erläutert.

1. INTRODUCTION

Requirements for forces engaged in peacekeeping and anti-terror missions demand reconnaissance systems with all-weather capability delivering in-time information content. A solution to these requirements are lightweight SAR-Systems on-board of RPVs. For the realization of compact and robust

systems with all-weather capability, millimeterwave radars are the prime choice. In addition to the compactness of the millimetric hardware, for which even integrated subsystems on GaAs-basis are feasible, very simple and straight-forward signal-processing techniques can be used at millimeterwave frequencies. This is mainly due to three reasons (Boehmsdorff, S., 1989): The aperture length of the SAR is reduced by the

ratio of wavelengths at equal cross-range resolution, and allows lower requirements for the flight stability of the airborne platform. The amount of excessive backscattering at edges and corners of man made structures is much less due to a relatively higher geometric deviation of right angles and straightness of lines and a higher roughness of surfaces. Interferometric SAR at millimeterwave frequencies offers the additional advantage of small baselines for high height estimation accuracy. However the smaller wavelength leads also to a higher phase ambiguity, resulting in a much higher amount of sophistication for phase unwrapping at these radar bands.

To overcome this shortcoming, a hardware approach has been developed using multiple baselines to deliver interferograms at different height estimation precision ranging from low to high. This approach is adequate for millimeterwaves as those different baselines can be accommodated within small geometric dimensions. For classical radar bands this can not be achieved within dimensions typical for RPVs.

2.A MULTIPLE BASELINE APPROACH FOR MILLIMETERWAVE INTERFEROMETRIC SAR

2.1 Hardware Assembly

The millimeterwave radar system for which the algorithm has been developed (Boehmsdorff, S., 2001) is equipped with a multibaseline antenna consisting of an array of six horns followed by a cylindrical lens, which achieves the azimuthal beamforming. The antenna has a 3 dB beam width of 3° in azimuth and 12° in elevation. A photo of the antenna arrangement is shown in Fig 1. E1 – E4 denote the four receiving channels, S1 and S2 the two possible transmitting ports.



Figure 1: Radar Front - End with Multiple Baseline Antenna

From each receiver channel complex data are retrieved. Due to the geometry of the horn ensemble five independent interferograms can be extracted.

2.2 Basic Phase Unwrapping

The phase unwrapping method based upon this hardware approach needs at least one unambiguous interferogram, which is usually the one corresponding to the smallest baseline. In case that this one is not unambiguous, it is necessary to unwrap it with an algorithm based upon a different approach. This can

be a quite simple one like the Goldstein or Dipole algorithm [2], if the height variation of the imaged area doesn't exceed the unambiguous height dramatically. In order to start the further processing all available interferograms are stored and sorted according to their baselines. The full process is tested using a synthetic scene with a number of generic objects. This is described in the following.

3.DESCRPTION OF SYNTHETIC TEST AREA

The test area described here is an artificial data set, created by a program written in Matlab®. Its dimension is 1000×2000 pixels, and it contains several sample objects of different heights and shapes. In particular there are seven generic objects to test the unwrapping algorithm and in addition two noise patches. All objects are located on a flat plane. A picture of the elevation model, which is described qualitatively in the following, is shown in Fig. 2.

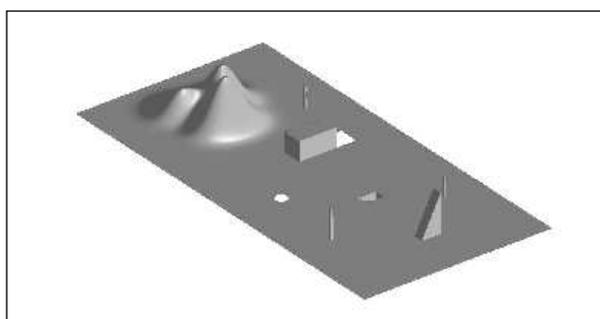


Figure 2: Elevation model of the test area. Noise patches are shown as white holes in the plane.

The first object at the upper left corner is a set of three mountains of Gaussian shape and different height and width neighbored by a round pillar. Below these there is a box, representing a man made object like a building, with a noise patch in the back simulating the shadow of the object. The next object is a round noise patch resembling a water surface where the backscattering tends to be of specular character and the phase appears noisy. Then a ramp is leading below the plane like a pit. Two rectangular pillars of different heights and an increasing ramp above a rectangular cross section complete the synthetic test area. For this test area the effectiveness of a new phase unwrapping technique, based upon an approach using more than two receiver channels at increasing interferometric basewidth will be evaluated.

4.DISCUSSION OF BASIC INTERFEROGRAMS

From this elevation model five interferograms were calculated for a baseline ratio of 1:2:3:4:5 relative to the first one. For the test it is assumed that the reference interferogram is unambiguous. All other interferograms are allowed to exhibit as many phase excursions as appropriate for the respective height. Due to the baseline ratios they are limited to five. In addition all simulated interferograms are corrupted with white noise of $\pm 15^\circ$ to resemble a typical measured data set. The phase values corresponding to the ground-plane are allowed to exhibit an arbitrary offset relative to each other. The five interferograms are shown in Figs. 3-7. To localize the noise patches the pseudo correlation map is rather used. The pseudo correlation map (Ghiglia, D. C., 1998) can be used as a quality criterion when

the magnitudes of the complex values are unknown. It is defined by:

$$|z_{m,n}| = \frac{\sqrt{(\sum \cos \varphi_{i,j})^2 + (\sum \sin \varphi_{i,j})^2}}{k^2} \quad (1)$$

The two sums are evaluated in an interval of the dimension $k \times k$ around each pixel (m,n) . The values of $\varphi_{i,j}$ are the phase values of the interferograms. Therefore this quality map is always available, even if only the phase differences are known. Within the correlation map points of low correlation indicate areas where the phase is decorrelated. These regions are noisy patches, like shadow areas or any region which shows specular reflectance. As an example Fig. 8 depicts the pseudo correlation map for the interferogram with the widest baseline shown in Fig. 7.

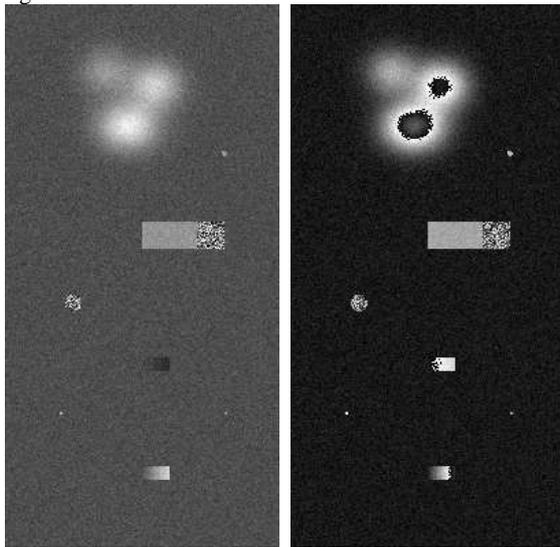


Figure 3: Interferogram, Baseline 1

Figure 4: as Fig. 3, Baseline 2

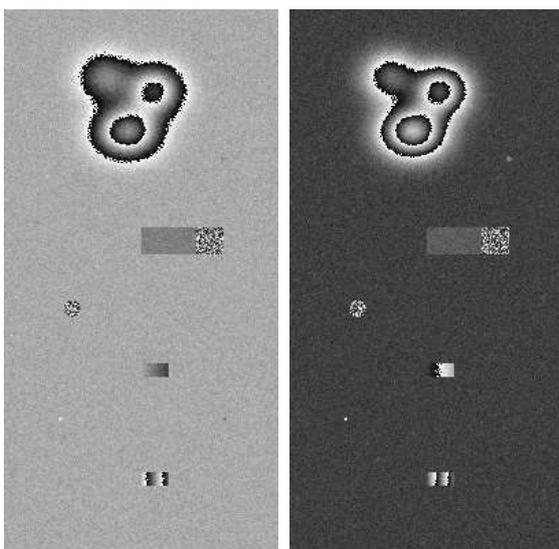


Figure 5: as Fig. 3, Baseline 3

Figure 6: as Fig. 3, Baseline 4

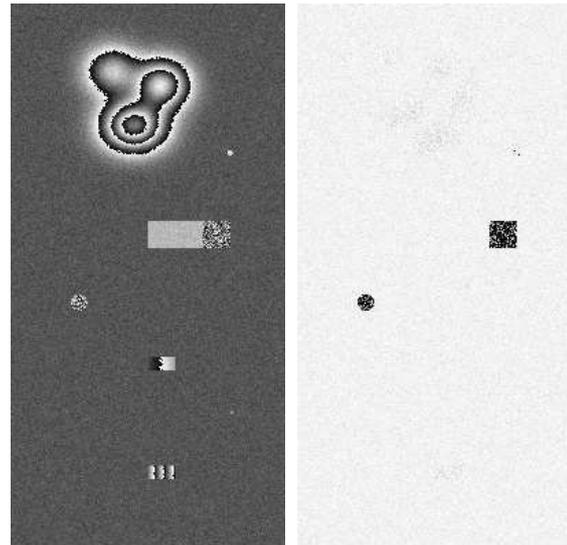


Figure 7: as Fig. 3, Baseline 5

Figure 8: Pseudo correlation map related to Fig. 7

The chart of Fig. 8 shows clearly the areas of low correlation within the noisy patches. In addition the limits of the box, the two pillars and the descending ramp are marked as areas of low correlation because of the considerable phase variation, although the phase within that limits is well defined. The border of the ascending ramp is marked with dashed lines which is due to the periodic correlation of the phase values (modulo 2π) while the phase values are increasing in comparison to that of the ground plane. Although the pseudo correlation map has the disadvantage of indicating steep terrain as low correlated it can be used as a tool to guide the phase unwrapping algorithm. In the example it serves to define areas where phase unwrapping is possible. After creating the interferograms at increasing baseline width it is possible to establish the unwrapping algorithm.

5. PHASE UNWRAPPING ALGORITHM FOR MULTIPLE BASELINE INSAR DATA

5.1 Description of the Process

As mentioned under 2.2 the algorithm needs at least one unambiguous interferogram as an initial starting point. It is postulated for the evaluation process, that the first interferogram, related to the narrowest baseline, is unambiguous. For further processing all interferograms are sorted according to their baselines. The unwrapping algorithm works iteratively using one interferogram after the other. From the preceding interferogram a look-up table is generated and used for unwrapping the next interferogram with higher interferometric base width where the phase is ambiguous but exhibits a higher precision. This is repeated until all interferograms are processed successively. It has to be pointed out, that unlike in the usual residues approach, no paths around ambiguous areas have to be searched to unwrap the phase. All undefined pixels within the resulting interferogram are strongly restricted to their location, and there is no error spread in the image plane. Errors that occur are mainly due to excessive phase noise. Due to the scaling of the successive interferograms which has to be done during the unwrapping process, the noise contributions are also scaled by the same ratio and therefore

erroneous pixels may occur. To solve this problem, the algorithm searches for suspicious isolated pixels within the resultant interferogram. For those pixels a test procedure determines if adding $\pm 2\pi$ gives a better result matching them to the surrounding pixels. The flow chart of the unwrapping algorithm is given in Fig. 9.

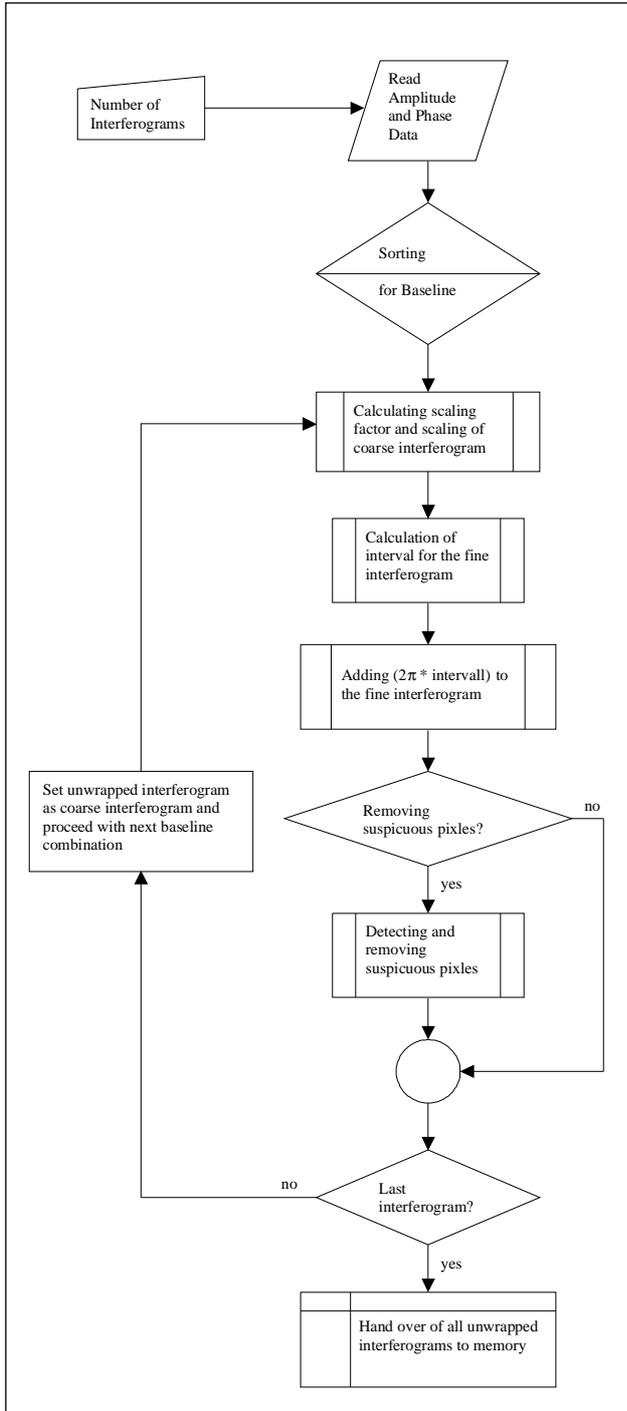


Figure 9: Flow chart of the unwrapping algorithm

Within all figures the gray scale is the same, where black represents $-\pi$ and white π . The algorithm starts with the first channel combination and unwraps all following ones. To visualize the success of the coarse to fine unwrapping process Fig. 10 shows the raw phase as vertical profile related to the synthetic area and Fig. 11 the corresponding unwrapped phase. It can be seen that the algorithm leaves most values untouched except at those positions where the phase is wrapped around. It has to be pointed out, that the algorithm presented here works fully automatically after choosing the input parameters and provides the user with five interferograms of different quality that have no or only a few residues left outside the noisy patches of shadow areas.

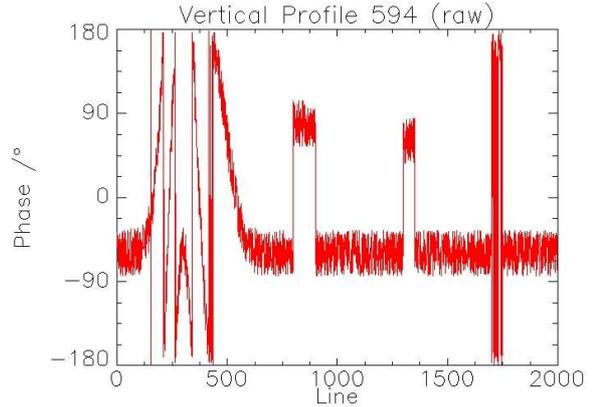


Figure 10: Vertical Profile of raw Phases

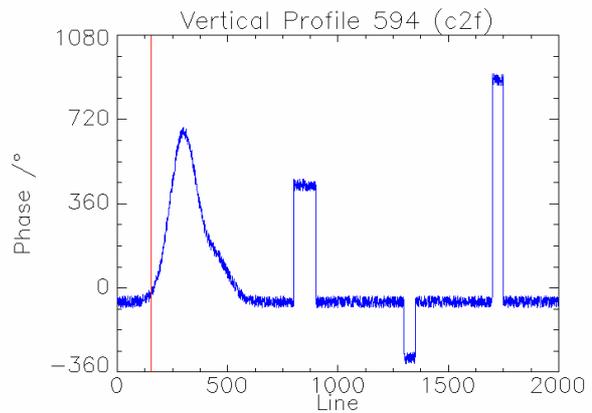


Figure 11: Vertical Profile gained by Coarse to Fine Process

5.2 Comparison with a Conventional Unwrapping Method

To demonstrate the advantage of the new coarse to fine unwrapping process (C2F) a conventional unwrapping method using residues has been applied to the model scene. Fig. 12 shows the comparison. It is obvious, that at those parts of the interferogram, where noise patches or even singular noise pixels appear the conventional method runs into problems and errors are spreading into the interferogram. To eliminate them a special treatment has to be done. The C2F-method does not suffer from such difficulties. As demonstrated under chapter 5.1 a vertical profile through the interferogram plane has been calculated and is shown in Fig. 13. It shows the difficulties for the Dipole method to cope with certain features of the

interferogram, namely holes in the plane and increasing ramps with multiple phase periods. Comparing the calculation time requirements it turns out, that the C2F Algorithm is about ten times faster than the conventional process.

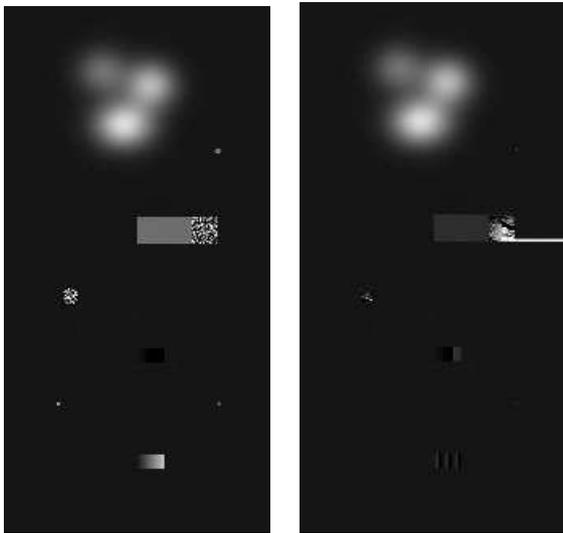


Figure 12: Comparison of Results of C2F-process and conventional Dipole Method

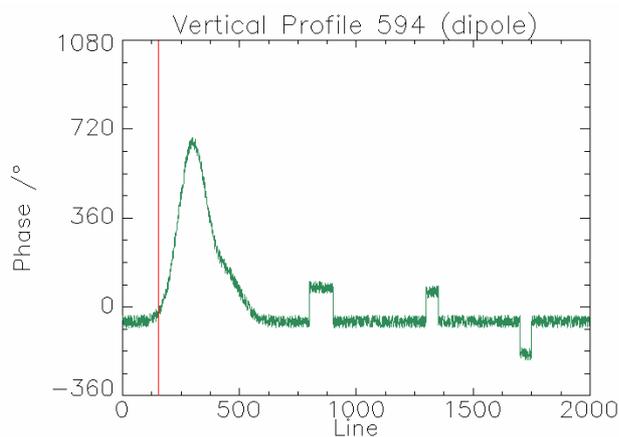


Figure 13: Vertical Profile gained by Dipole Method

Fig. 14 shows a section of the residue chart for the circular region incorporating phase noise contributions in comparison with the corresponding section of the way chart. The latter describes the path which may not be exceeded by the phase unwrapping process.

The example shows, that even with multiple folding of the interval $[-\pi, \pi]$ the algorithm delivers correct values, as long as the interferogram is unambiguous. The method avoids a time consuming calculation of the adequate integration path with the expense of a higher amount of hardware. It has further to be pointed out, that there is no error continuation within the image plane as with residue based methods, and all erroneous values stay strictly limited to its vicinity.

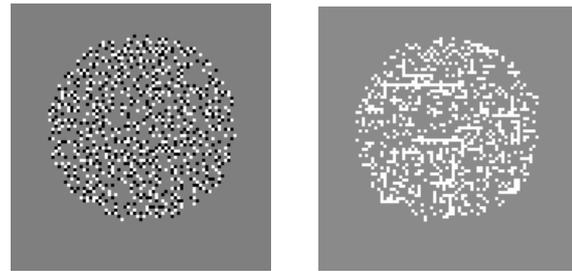


Figure 14: (left) Residue Chart, negative Values = white, positive Values = black, (right) Representation of the Way Chart, allowed Ways = grey, Not allowed Ways = white

6.SUMMARY

At mm-wave frequencies even with moderate baselines, i.e. in the order of 30 cm, a good height estimation is possible, which is appropriate to resolve the shape of man-made objects. On the other hand a severe disadvantage is the high height ambiguity due to the short phase repetition frequency at mm-wavelengths. So much emphasis has to be put into the phase-unwrapping algorithm. Appropriate for the possibility of small geometrical dimensions and ease of processing is the hardware approach of a multiple baseline antenna which makes the employment of the described coarse-to-fine algorithm possible. Because scaling is important for this process an appropriate ratio of consecutive baselines is essential. Good results can be achieved with a ratio of 5. Nevertheless the limitation of this algorithm is, that one initial unambiguous interferogram is necessary to start with. As the absolute phase variation in each unwrapped interferogram is constant, the relative phase variation becomes smaller within each of the consecutively unwrapped interferograms. The system and the algorithm presented here, is a suitable approach for mm-wave frequencies where the dimension of the multibaseline antenna itself does not exceed that of the sensor front-end.

7.REFERENCES

- Boehmsdorff, S., 1989. MEMPHIS an Experimental Platform for Millimeterwave Radar, DGON Intl. Conf. on Radar, München 1998, pp. 405-411
- Boehmsdorff, S., 2001. Detection of Urban Areas in Multispectral Data, IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas, 2001
- Ghiglia, D. C., 1998. *Two-Dimensional Phase Unwrapping*, John Wiley & Sons, Inc. New York, 1998

A “SIGHT-SPEED” HUMAN-COMPUTER INTERACTION FOR AUGMENTED GEOSPATIAL DATA ACQUISITION AND PROCESSING SYSTEMS

Gennady Gienko^{a*}, Eugene Levin^b

^aThe University of the South Pacific, Fiji Islands - gennady.gienko@usp.ac.fj

^bMichigan Technological University, Houghton MI, USA – e Levin@mtu.edu

KEY WORDS: eye-tracking, human-computer symbiosis, photogrammetry, 3D modelling, knowledge elicitation

ABSTRACT

Many real-time tasks in geospatial data analysis are based on matching of visual data, i.e. finding similarity and/or disparity in geoinages either in the remotely sensed source data or in geospatial vector and raster products. When human eyes scrutinize a scene, the brain performs matching of visual streams, acquired by eyes and transmitted via all chains of human visual system. As the result the brain creates a comfortable visual model of the scene and alarms, in case some distinct disparity in visual perception is found. While observing a scene, the optical axes of both eyes are naturally directed to the same area of the object, which is particularly true for visual perception of stereoscopic images on a computer screen. If eye movements are recorded while observing the virtual stereoscopic model generated in the brain, it is possible to detect such regions of interest using fixations, identified in eye-tracking protocols. These fixations can be considered as coordinates of feature points of an object being observed (regions of interest) and can be used to reconstruct corresponding 3D geometric models by applying classical stereo photogrammetric procedures. This novel way of utilizing eye-tracking leads to establishment of “eye-grammetry” - a new approach which melds human visual abilities and the computational power of modern computers to provide “sight-speed” interaction between the human operator and the computer software in augmented geospatial data acquisition and processing systems. This paper reviews theoretical and practical aspects of eye-tracking for real-time visual data processing and outlines two particular fields of applications where the proposed technology could be useful: gaze-tracking based 3D modeling and geospatial knowledge elicitation.

1. INTRODUCTION

Many real-time tasks in geospatial data analysis are based on matching of visual data, i.e. finding similarity and/or disparity in geoinages either in the remotely sensed source data or in geospatial vector and raster products. Image fusion, change detection, 3D surface reconstruction, geospatial data conflation, – these are the only few examples of tasks that employ visual matching. Humans can instantaneously “sense” visual disparity due to the fundamental capabilities of the human visual system to perform matching. When human eyes observe a scene, the brain performs matching of visual streams, acquired by eyes and transmitted via all chains of the human visual system.

The brain creates a comfortable visual model of the scene and alarms, if some distinct disparity in visual perception has been found. Human-computer symbiosis, in augmented geospatial data acquisition and processing systems is based on eye-tracking techniques that makes it possible to arrange a “sight-speed” loop for interaction of human operator and computer software.

The virtual scene, imagined in the brain, is inherently related to neuro-physiological features of the human visual system and differs from the real world. The brain processes visual input by concentrating on specific components of the entire sensory area so that the interesting features of a scene may be examined with greater attention to detail than peripheral stimuli. Visual attention, responsible for regulating sensory information to sensory channels of limited capacity, serves as a “selective filter”, interrupting the continuous process of ocular observations by visual fixations. That is, human vision

is a piecemeal process relying on the perceptual integration of small regions of interest (ROI) to construct a coherent representation of the whole.

While observing a scene, the optical axes of both human eyes are naturally directed to the same area of the object, which is particularly true for visual perception of stereoscopic images on a computer screen. Human eyes, under subconscious control, move very rapidly to scan images and the result of this scanning is sent to the brain.

If eye movements are recorded while observing the virtual stereoscopic model, generated in the brain, it is possible to detect such regions of interest using fixations, identified in eye-tracking protocols. These fixations can be considered as coordinates of feature points of an object being observed (regions of interest) and can be used to reconstruct corresponding 3D geometric model, applying classical stereo photogrammetric procedures. This novel way of utilizing eye-tracking data leads to establishment of “eye-grammetry” - a new branch of photogrammetry, which synthesizes human visual abilities and fundamentals of classic stereometry for real-time 3D measurements.

While it is generally agreed upon that fixations correspond to the image measurements, it is less clear exactly when fixations start and when they end. Common analysis metrics include fixation or gaze durations, saccadic velocities, saccadic amplitudes, and various transition-based parameters between fixations and/or regions of interest. The analysis of fixations and saccades requires some form of fixation identification (or simply identification) - that is, the translation from raw eye-movement data points to fixation

* Corresponding author

locations (and implicitly the saccades between them) on the visual display. Fixation identification is an inherently statistical description of observed eye movement behaviors. Comparative study of fixation identification algorithms (Salvucci and Goldberg, 2002) suggests dispersion-threshold method as a fast and robust mechanism for identification of fixations. This method is also quite reliable in applications requiring real time data analysis, which is a critical aspect in real-time photogrammetric applications.

The previous research outline theoretical and practical aspects of combining the human capability of matching visual images with a computer's capability of fast calculation, data capturing and storage to create a robust system for real-time visual data processing (Gienko and Chekalin, 2004). Theoretical research has been done to investigate neuro-physical features of human visual system and analyze technical parameters of eye-tracking systems. Research work was aimed on designing of a prototype of the system, developing algorithms for stereoscopic eye-tracking and investigation of accuracy issues of visual perception and stereoscopic measurement of geospatial information, - aerial photographs in particular (Gienko and Chekalin, 2004, Gienko and Levin, 2005).

The present paper describes two fields of applications where the proposed technology could be useful: 1) fast generation of 3D models based on eye movement measurements during observation by human operators of stereoscopic models; 2) knowledge elicitation: automated eye-tracking allows establishment of a protocol of an expert's conscious and subconscious processes during visual image interpretation sessions, that enables extraction and formulation of knowledge which, being asked, experts are usually unable to articulate.

2. VISUAL PERCEPTION: SEEING AND MATCHING

Typical tasks in geospatial data visual analysis include, but not limited to retrieval of information, image interpretation, change detection, 3D surface reconstruction and updating of derived geospatial data such as GIS vector layers. In many application scenarios such as risk management or military targeting etc. it is required to perform these tasks in the real-time mode. Specifically all these tasks require visual data matching and fusing performed by a human analyst, who at the same time can be a Subject Matter Expert (SME) and, under certain circumstances act as a Decision Maker. Thus, the solutions described below constitute some useful technology empowering certain types of decision support systems, which in terms of Computer Sciences can be defined as a Human-Computer Symbiosis (HCS) in visual data analysis.

Table 1 outlines main stages of a typical image analysis process which usually involves certain human intellectual and computerized recourses, employed simultaneously or concurrently, whichever is the most effective for a particular task:

Table 1. Human and computers in image analysis

Stage	Agent	
General matching of observed scenes	brain	
Tuned area matching	brain	computer
Disparity evaluation	brain	computer
Finding spot correspondence	brain	
Object recognition	brain	
Measuring (un)matched objects	brain	computer
Measurements registration		computer
Statistics		computer
Analysis	brain	computer

This authors' point of view on comparative effectiveness of human analysts and automated computer programs at a particular stage of image analysis prompts us to develop a human-in-the-loop technology for processing of geospatial visual data in the most efficient way. As humans perceive and process vast amount of information about the environment through their visual system at extremely high speed, it is seems reasonable to combine this human's ability and computational power of computers to build a Human-Computer Symbiosis platform for processing of visual geospatial data. Such HCS can be based on registering of visual activity of an operator using techniques of real-time eye-tracking. While the human brain performs searches and analysis of visual data, operator's eyes subconsciously scan the visual scene. Such eye movements are driven by and indirectly represent results of internal processes of visual searching and matching, performed by the whole human visual system. Tracking and analyzing eye movements allows us to arrange a 'sight-speed' loop with the computer which should perform the rest of the tasks where computations and data storage are predominant.

3. VISUAL PERCEPTION AND EYE MOVEMENTS

The virtual scene, imagined in the brain, is inherently related to neuro-physiological features of human visual system and differs from the real world. The brain processes visual input by concentrating on specific components of the entire sensory area so that the interesting features of a scene may be examined with greater attention to detail than peripheral stimuli. Visual attention, responsible for regulating sensory information to sensory channels of limited capacity, serves as a "selective filter", interrupting the continuous process of ocular observations by visual fixations. That is, human vision is a piecemeal process relying on the perceptual integration of small regions to construct a coherent representation of the whole.

Neurophysiological literature on the human visual system suggests the field of view is observed through brief fixations over small regions of interest (ROIs) (Just and Carpenter, 1984). This allows perception of detail through the fovea. When visual attention is directed to a new area, fast eye movements (saccades) reposition the fovea. Foveal vision allows fine scrutiny of approximately 3% of the field of view but takes approximately 90% of viewing time, when subtending 5 deg of visual angle occurs. A common goal of eye movement analysis is the detection of fixations in the eye

movement signal over the given stimulus or within stimulus ROIs.

It has been found (Mishkin et al., 1983) that humans and higher animals represent visual information in at least two important subsystems: the where- and the what systems. The where-system only processes the location of the object in the scene. It does not represent the kind of object, but this is the task of the what-system. The two systems work independently of each other and never converge to one common representation (Goldman-Rakic, 1993). Physiologically, they are separated throughout the entire cortical process of visual analysis.

When the brain processes a visual scene, some of the elements of the scene are put in focus by various attention mechanisms (Posner et al., 1990). When the brain analyses a visual scene, it must combine the representations obtained from different domains. Since information about the form and other features of particular objects can be obtained only when the object is foveated, different objects can be attended to only through saccadic movements of the eye – the rapid eye movements, which are made at the rate of about three per second, orienting the high-acuity foveal region of the eye over targets of interest in a visual scene. The characteristic properties of saccadic eye movements (or saccades) have been well studied (Carpenter, 1988).

Saccades are naturally linked with fixations – relatively stable positions of the eye during a certain time. Varieties of researches prove that visual and cognitive processing do occur during fixations (Just and Carpenter, 1984). The process of fixation identification is an inherently statistical description of observed eye movement behaviors and separating and labeling fixations and saccades in eye-tracking protocols is an essential part of eye-movement data analysis (Salvucci and Goldberg, 2000).

4. FROM EYE FIXATIONS TO IMAGE MEASUREMENTS

In continuous movements eyes can be relatively stable only limited time, in most cases 200 to 800 msec. These fixations in eye positions occur in and correspond to certain regions of interest where the eyes perceive featured objects of the scene or part thereof. Projection of a certain fixation into the object's plane corresponds to a gaze position which in case of an image displayed on computer screen corresponds to certain area of an image matrix which allows as consider these gaze positions as image measurements.

While it is generally agreed upon that fixations (through their projected coordinates into the object's plane) correspond to coordinates of points in image, it is less clear exactly when fixations start and when they end. Common analysis metrics include fixation or gaze durations, saccadic velocities, saccadic amplitudes, and various transition-based parameters between fixations and/or regions of interest (Salvucci and Goldberg, 2000). The analysis of fixations and saccades requires some form of fixation identification (or simply identification)—that is, the translation from raw eye-movement data points to fixation locations (and implicitly the saccades between them) on the visual display. Fixation identification is an inherently statistical description of observed eye movement behaviors.

For spatial characteristics, (Salvucci and Goldberg, 2000) identify three criteria that distinguish three primary types of algorithms: velocity-based, dispersion-based, and area-based. Velocity-based algorithms emphasize the velocity information in the eye-tracking protocols, taking advantage of the fact that fixation points have low velocities and saccade points have high velocities. Dispersion-based algorithms emphasize the dispersion (i.e., spread distance) of fixation points, under the assumption that fixation points generally occur near one another. Area-based algorithms identify points within given areas of interest (AOIs) that represent relevant visual targets. These algorithms provide both lower-level identification and higher-level assignment of fixations to AOIs. Because fixations can also be used as inputs to AOI algorithms, these can also represent higher levels of attentional focus on a display (Scott and Findlay, 1993). These dwell times can be considered 'macro-fixations', in that they organize fixations into a larger picture. For temporal characteristics, (Salvucci and Goldberg, 2000) include two criteria: whether the algorithm uses duration information, and whether the algorithm is locally adaptive. The use of duration information is guided by the fact that fixations are rarely less than 100 ms and often in the range of 200-400 ms. The incorporation of local adaptivity allows the interpretation of a given data point to be influenced by the interpretation of temporally adjacent points; this is useful, for instance, to compensate for differences between 'steady-eyed' individuals and those who show large and frequent eye movements.

Comparative study of fixation identification algorithms (Salvucci and Goldberg, 2002) suggests dispersion-threshold method as a fast and robust mechanism for identification of fixations. This method is also quite reliable in applications, requiring real time data analysis, which is a critical aspect in real-time photogrammetric applications (Gienko and Chekalin, 2004).

5. EYE-GRAMMETRY

Spatial and temporal data derived from eye movements, compiled while the operator observes the geospatial imagery, retain meaningful information that can be successfully utilized in image analysis and augmented photogrammetry. We call this technology Eye-grammetry - a new approach to a 'sight-speed' acquisition and processing geospatial visual data using real-time eye-tracking technologies. Eye-grammetry is derived from words "eye" and "grammetry" (measure) and stands for obtaining reliable information about physical objects and the environment by detection and analysis of human eye movements observing these objects or their visual representations in images.

In general, the word grammetry refers to non-contact measurements of the object from images. Nowadays we use a number of "grammetric" techniques, aimed on precise measurement of the object, pictured in images. To acquire these images, some very advanced technologies are used – in different spectral zones and data presentations. Every new technological break-through, resulting in appearance of a new sensor, introduces a new definition – radar-grammetry, sonar-grammetry, x-ray-grammetry, etc. Sometimes looking at some of such images it is hardly to say that it is an image in the sense, that it was used early last century for conventional photographs.

Several attempts have been made to introduce the broader definitions such as "iconoactinometry" to describe new methods of registration and visual representation of the imaged objects using modern techniques, but the term "grammetry" is still well known and widely accepted within the professional community. So, to keep the traditions, we name our method Eye-grammetry – a new technology for measuring and interpretation the images.

In general, eye-grammetry could be defined as a technology of obtaining quantitative information about physical objects and the environment. This is done through measuring and interpreting images, acquired by different terrestrial, airborne or satellite sensors. In contrast to traditional principles of creation of photogrammetric terms, the first word component introduces spectral characteristics of registered radiation (photo, radar, x-ray), the word eye in our definition is interpreted as a "tool" and grammetry is widened for "image measurements". Therefore, eye-grammetry means measuring of objects in images by the eyes.

Technically, eye-grammetry is a technology based on principles of tracking the human eye movements while perceiving the visual scene. Spatial and temporal data derived from eye movements, compiled while the operator observes geospatial imagery, retain meaningful information that was successfully utilized in image analysis and augmented photogrammetry. This challenge is achievable based on human stereopsis principles.

Human stereopsis declares that while observing a scene, optical axis of the both human eyes are naturally directed to the same area of the object, which is particularly true for visual perceiving of stereoscopic images on a computer screen. Processing recorded movements of eyes, subconsciously scanning scene or image, it is quite possible to identify centers of gravity of fixation areas, which correspond to (and coincide with) identical points of the object on the left and right images of a stereopair (Figure 1).

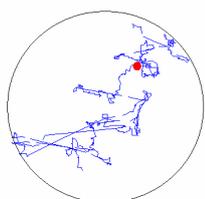


Figure 1. Eye movement trajectory in fixation area

Projected gaze directions of the operator's eyes for corresponding fixations can be interpreted as coordinates of the featured points of an object being observed in stereo image (Figure 2). Thus, the main challenge in eye-grammetry is identification of fixations in eye-tracking protocols and calculation corresponding gaze directions to define coordinates of points in observed images, which then can be treated as conventional photogrammetric measurements.

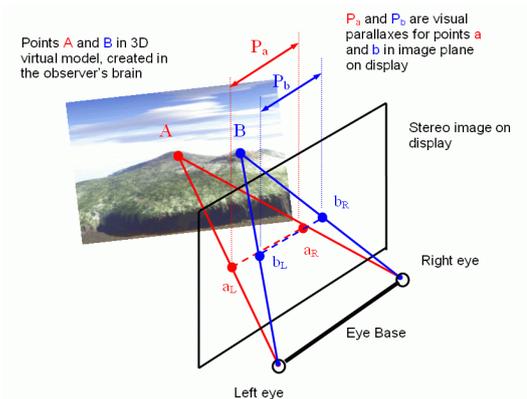


Figure 2. Stereoscopy in eye-grammetry

6. EYE TRACKING FOR 3D GEOSPATIAL MODELING

Several techniques are used to track eye movements. The electro-oculography technique is based on electric skin potential, and uses the electrostatic field that rotates along with the eye. By recording quite small differences in the skin potential around the eye, the position of the eye can be detected (Mowrer et al, 1936, Gips et al, 1993). If the users wear a special contact lens, it is possible to make quite accurate recordings of the direction of gaze. By engraving one or more plane mirror surfaces on the lens, rejections of light beams can be used to calculate the position of the eye (Ward, 2001). The most common techniques of eye tracking are based on rejected light. These techniques employ limbus tracking, pupil tracking and corneal reflection (Mauler et al., 1993, Ward, 2001). The highest spatial and temporal resolution could be achieved using the dual-Purkinje eye-trackers (Cornsweet and Crane, 1973).

Design of an eye-tracking system optimal for the geospatial data processing was an initial effort of the current research (Gienko and Chekalin, 2004, Gienko and Levin, 2005). Figures 3 and 4 demonstrate the principal design and working prototype of the system demonstrated in 2004 at XXth ISPRS congress in Istanbul (Geoiconics 2004, Intelligaze 2007).

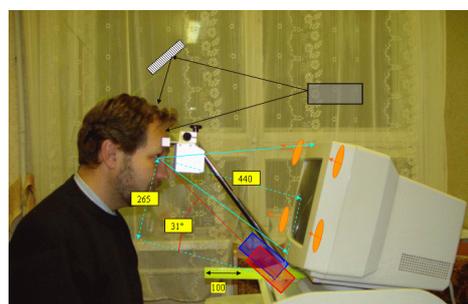


Figure 3. Principal design of eye-tracking system for geospatial visual data processing

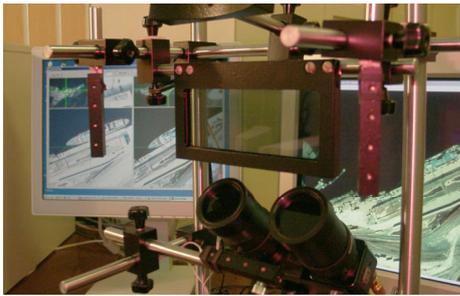


Figure 4. Working prototype of eye-tracker for processing of geospatial visual data

Calibration of precise eye-tracking systems is a bottleneck in augmented photogrammetric systems. Depending on chosen technique and hardware process of calibration involves the following major steps: 1) photometric calibration video cameras; 2) estimation positions of IR LEDs in order to estimate center of cornea; 3) resolving the geometric properties of a monitor; 4) determining relationship between video cameras and the screen to transform camera and monitor coordinate systems; and 5) determining the angle between visual-optical axis.

Once calibrated, the photogrammetric eye-tracking system can be used for two major applications, which involve matching of geospatial visual data: a) generation of 3D models based on eye-tracking; b) knowledge extraction based on eye-tracking protocols of the Subject Matter Experts (SME) and Decision Makers.

Figure 5 illustrates major stages of photogrammetric eye-tracking process for 3D modeling, assuming that eye-tracking system has been calibrated and the human analyst observes the scene stereoscopically under comfortable and stable stereoscopic conditions.

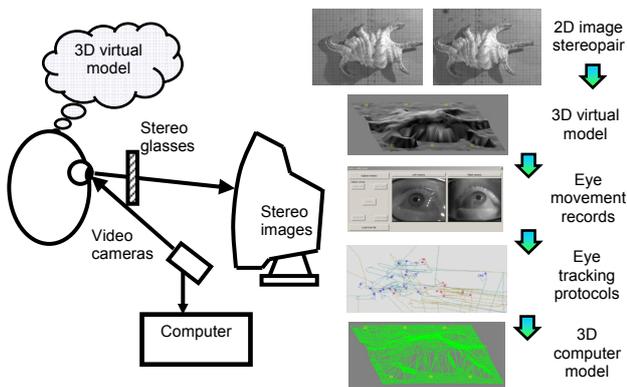


Figure 5. Principles of 3D scene restoration based on eye-movement measurements

The challenge in this technology is to extract a set of discrete and well-defined image measurements to reconstruct 3D model of a scene in the real-time. Point measurements are derived from fixations, which, in turn, statistically calculated and extracted from eye-movement protocols using individually set criteria, defined as a result of geometric calibration which contains personal data and parameters for each human analyst. Calibration process is personalized and sensitive to physiological parameters of eyes. Apart from projection parameters to calculate gaze position on the

screen, the calibration involves compensation of head movements which is the second derivative and partially correlated with saccadic eye movements.

7. EYE-TRACKING AND GEOSPATIAL SME KNOWLEDGE EXTRACTION

The idea of using eye-tracking for geospatial Subject Matter Expert (SME) knowledge elicitation is based on discovering and formalizing associations and correlations between content of the image observed, expert's eye-movements trajectories and particular tasks given to an expert – whether it is targeting of specific objects in a set of multi-sensor and multi-temporal images, pure image classification or some other task involving geospatial data such as maps and GIS layers or other visual information. The system tracks the expert's gaze directions while he selects and labels objects, then calculates parameters of the selected objects and generates preliminary classification rules by applying a dedicated knowledge mining approach.

The challenge of this approach is to improve data mining procedures by means of the rules extracted from human analyst deploying eye-tracking system. Technological scheme of the eye-tracking based visual knowledge extraction is depicted in Figure 6.

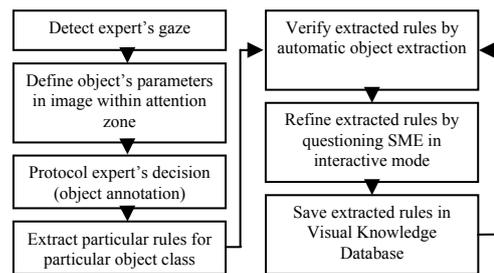


Figure 6. Eye-tracking based knowledge elicitation process

Once the expert finishes natural process of image recognition, the full set of extracted rules is verified by re-applying those rules by automatic classification of the same source image. All automatically extracted and classified objects then matched against the results of the expert's natural work. Unmatched objects indicate inadequacy of the extraction rules. The expert interactively reviews and verifies results of image interpretation to discover the reasons of inadequacy which then will be used to adjust algorithms and parameters for automated extraction of decision rules. It is an iterative and interactive process, so the results will be immediately applied to the source image and the expert will be able to evaluate effectiveness of the newly added or modified rule. Once finished, the system will apply "reverse rule" verification to cluster extracted rules and rate them in order to select the minimum set of major rules and knowledge sufficient for robust classification of particular objects.

The system is designed to implement self-learning concept to accrue results of classification of the same image carried out by a number of experts with different levels of expertise. The system allows Subject Matter Experts (SMEs) to formalize and transfer their imagery knowledge into knowledge-based reasoning systems most efficiently, with minimal help of

knowledge engineers. Conceptually this technology is based on research in neurophysiologic features of human visual system (HVS), particularly related to Gestalt rules, and cognitive associations while perceiving meaningful visual information.

8. CONCLUSIONS AND OUTLOOK

Eye-grammetry is a very new direction in geospatial data acquisition, processing and analysis. Based on eye-tracking methods, eye-grammetry synthesizes human visual abilities and computational power of computers to build a new kind of Human-Computer Symbiosis, specifically designed to solve variety of tasks that involves extensive processing of geospatial visual data – from measurements to object recognition. Applications of eye-grammetry in geospatial technologies can be numerous – 3D modelling and eye-guided selective LIDAR data cleaning, DEM compilation and interactive geodatabase updating using visual data fusion, natural disaster assessment and decision making support in Geographic Expert Systems, education, training and Real-Time Expertise Transfer (RTET), air-traffic control and geo-monitoring and warning systems, homeland security and surveillance, etc.

Further theoretical and practical research and investigations should be carried out towards comprehensive analysis of neuro-physiological features of human visual system, particularly on study of optical and physical eye parameters for observations of 3D virtual models by viewing stereo images in photogrammetric applications; precision and accuracy issues - such researches should encompass an impact of digital image resolution, video frames frequency and visual a-synchronism of left/right eyes on the accuracy of identification of fixations. Developing rigorous mathematical models to link light-eye-camera-object parameters for precise geometric calibration is another niche for extensive investigations. Hardware and optical limitations, real-time hires video stream processing are the other challenges – some alternative programming languages and approaches have to be considered to ensure effectiveness of image measurements and data processing.

REFERENCES

- Carpenter R.H.S., (1988). *Movements of the Eyes*. London: Pion.
- Cornsweet, T. N., Crane, H. D., (1973). US Patent US3724932.
- Geoiconics (2004). Corporate materials, <http://www.geoiconics.com>
- Gienko, G., and Chekalin, V. (2004). Neurophysiological features of human visual system in augmented photogrammetric technologies. Proc. XX ISPRS Congress, Istanbul, 12-23 July, 2004.
- Gienko, G., and Levin, E., (2005). Eye-tracking in augmented photogrammetric technologies. Proc. of ASPRS Int. conference, Baltimore, USA, March 2005
- Gips, J. and P. Olivieri, and J. Tecce. (1993). Direct control of the computer through electrodes placed around the eyes. IEEE Transactions on Systems, Man, and Cybernetics, 20(4):630-635, 1993. In Fifth International Conference on Human-Computer Interaction, (HCI International '93).
- Goldman-Rakic, R.S., (1993). "Dissociation of object and spatial processing domains in primate prefrontal cortex", Science, 260, 1955-1957.
- IntelliGaze (2007). Corporate materials, <http://www.intelligaze.com>
- Just, M.A., and Carpenter, P.A., (1984). Using eye fixations to study reading comprehension. In D. E. Kieras & M. A. Just (Eds.), *New Methods in Reading Comprehension Research* (pp. 151-182). Hillsdale, NJ: Erlbaum.
- Muller, P.U., D. Cavegn, G. d'Ydewalle, and R. Groner (1993). A comparison of a new limbus tracker, corneal reflection technique, purkinje eye tracking and electro-oculography. In G. d'Ydewalle and J. V. Rensbergen, editors, *Perception and Cognition - Advances in Eye Movement Research*, pages 393-401. Elsevier Science Publishers.
- Mishkin M., Ungerleider, L. G. and Macko, K. A., (1983). Object vision and spatial vision: Two cortical pathways, *Trends in Neuroscience*, 6, 414-417.
- Mowrer O.H., and T.C. Ruch, and N.E. Miller. The corneo-retinal potential difference as basis of the galvanometric method of recording eye movements. *Am J Physiol*, 114:423-428, 1936.
- Posner, M.I., and Petersen, S.E., (1990). "The attention system of the human brain", *Annual Review of Neuroscience*, 13, 25-42.
- Salvucci, D. D., and Goldberg, J. H., (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the Eye Tracking Research and Applications Symposium* (pp. 71-78). New York: ACM Press.
- Scott, D. and J. M. Findlay, (1993). Visual search, eye movements and display units. Technical report, before 1993.
- Ward, D. J. (2001). *Adaptive Computer Interfaces*. PhD thesis, Churchill College, Cambridge.

DIGITAL SURFACE MODELS FROM ULTRACAM-X IMAGES

M. Gruber, S. Schneider

Microsoft Photogrammetry, Anzengruebergasse 8, A-8010 Graz - (michgrub, susannes)@microsoft.com

KEY WORDS: UltraCamX, Digital Surface Models, Digital Images, Large Scale Mapping, Aerial Images

ABSTRACT:

UltraCamX is the newest large format digital aerial camera system by Microsoft Photogrammetry (former Vexcel) and was introduced in 2006. We present a short technical overview about the sensor and we show results from a data experiment. Images from this sensor were produced at high overlaps and used to create digital surface models. In this paper the geometric accuracy of the digital surface data and the quality of UltraCamX images are presented.

UltraCamX offers to capture one frame at an interval of 1.35 seconds. This allows to producing images at high overlaps at large scale and reasonable speed of the aircraft. For the data experiment a ground sampling distance of 10cm was produced at a forward overlap of 90%. Such redundant image data can be used to improve automation and robustness during the processing of dense surface models.

1. INTRODUCTION

Overlapping images are one important source data set for three dimensional object reconstruction and surface modeling. This application - known as "shape from stereo" – does rely on the quality of the source images. Geometric accuracy and radiometric bandwidth are the two most important requirements which are requested in order to produce quality results. Since a few years large format digital aerial cameras are available and did find their way into the photogrammetric production. It was very obvious that the digital sensor did show advantages over film based, scanned images. The most important advantage can be seen in the radiometric quality, thus no longer film grain noise affects the matching process.

Another advantage of digital sensors can be seen in the new all digital workflow excluding the expensive and time consuming handling of the film. Digital image data can be produced and stored at very attractive costs. Therefore the acquisition of data capture may be reorganized. Not the minimum of images but redundancy and robustness are the requirements of the all digital photogrammetric projects.

The new digital sensor system by Microsoft, the UltraCam X was used to capture images over a well known project area. Digital surface information was extracted from UltraCamX images by means of image analysis. The project area of Gleisdorf was mapped at a ground sampling distance (GSD) of 10 cm at a special flight pattern of high overlap. 9 lines are flown in N-S direction, with each about 31 images and 5 lines in the E-W direction, with about 25 images. The project is flown with 80 % forward overlap and 65% side overlap.

We present the quality of the resulting Digital Surface Model (DSM) datasets. From one test flight one DSM was derived from the North to South flight lines and one from the East to West flight lines. The resulting DSMs are compared with Ground Control Points (GCP) in the area and z-Differences are shown.

1.1 UltraCam X

The UltraCam X large format digital aerial camera system was introduced in 2006 and can be seen as the successor of Vexcel's UltraCam D. The basic concept of the design did not change. The sensor unit consists of eight camera heads. Four camera heads are responsible for the large format panchromatic image, another set of four camera heads contribute with four color bands in red, green, blue and near infrared. The sensor produces images of 14430 by 9420 pixels at a frame rate of 1 frame in 1.35 seconds. Thus high overlaps of 80 % along track can be achieved at 5 cm GSD and a reasonable speed of the aircraft. The physical dimension of the image is 103.9 mm by 67.8 mm at a Pixel size of 7.2 μm . The principal distance of the camera is 100.5 mm.



Fig. 1: UltraCam X digital aerial camera system with the Sensor Unit (right) and the airborne Computing Unit including two removable Data Units (left).

The data storage system of the camera offers storage space for almost 4700 frames, thus the system is well prepared for intensive image harvest at large scales and at high overlaps. All image data are initially stored at 16 bit bandwidth (see Fig. 1).

1.2 Basic source data

During the routine delivery process every UltraCamX camera systems has to pass a specific test flight which is performed over the test field of Gleisdorf near Graz. Figure 2 shows an overview of the flight plan for that area. The size of the Gleisdorf test field is about 5 km by 7 km. It shows a huge diversity of cultivation including build up areas, traffic infrastructure, agricultural plans and forests. The ground truth consists of 56 full ground control points and a local GPS station.

The source data sets for our DSM quality assessment consist of images taken at a ground sampling distance of 10 cm at a flying height of 1400 m above ground level. The forward overlap was set to 80% the side lap was set to 65% resulting into a distance between flight lines of 500 m and a base length of 200 m. This special flight pattern supports redundancy and thus automatic blunder detection. Flight lines facing north to south and flight lines facing east to west were used to produce two different sets of DSMs.

During the flight mission GPS/IMU information was collected.

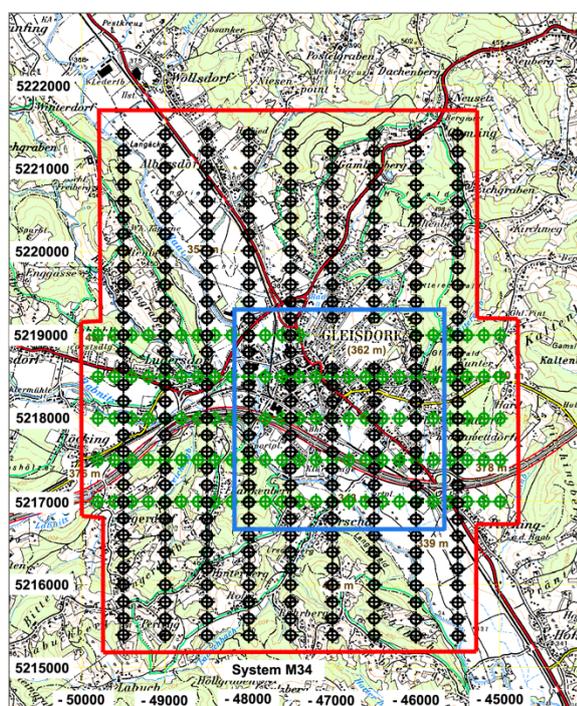


Fig. 2: Overview of the Gleisdorf area with flight lines in two directions and 500 m spacing. The base distance between photo centers is 200 m. The DSM area is marked up by the blue quadrangle.

The size of the source images of the UltraCamX camera system is 14430 pixels cross track and 9420 pixel along track, thus a footprint of 1443 m by 942 m is covered by one frame at a flying height of 1400 m. The single base distance between two images is 200 m, thus the resulting base to height ratio is only 1/7. Therefore a set of three or more images is used for every point measurement. This leads to a base to height ratio of 1/3.5, 1/2.3 or even 1/1.75 and increases the vertical quality and the robustness of the set.

Fig. 3 shows one single frame of the camera system. The sub area on the right gives an impression of the level of detail and the radiometric quality of the image.

The flight mission took place on May, 10th of 2007.



Fig. 3: Source image from UltraCamX at a format of 14430 by 9420 pixels. The entire frame is shown on the left, the detail on the right covers a footprint of 68 by 90 meter.

1.3 The workflow

The entire photogrammetric process consists of the three basic steps of source data acquisition including the post-processing of the raw image data, geo referencing of the images based on GPS/IMU supported aerial triangulation and DSM computation by means of dense image matching.

The post processing is done by the camera specific software package OPC (Office Processing Center) and converts the set of 13 raw sub images of each frame into one single large format panchromatic image and one four band color image. This process includes the so-called stitching between overlapping sub images and also introduces the parameters of the laboratory calibration data set. Such post processed and therefore geometrically and radiometrically corrected images are then introduced into the aero triangulation. The automatic atriangulation (AAT) is performed by using the Match AT software product from INPHO GmbH, Stuttgart, Germany. A camera specific bundle adjustment is then added by using the BINGO software product from GIP, Aalen, Germany. The GPS and IMU processing was supported by IGI mbH, Kreuztal, Germany.

After the images are geo referenced a refined AT is computed and matching points are extracted. These procedures give additional information on the accuracy of the AT, used for the DSM computation in the next step. The DSM algorithm was especially developed for processing UltraCam images. A hybrid multi-view matching algorithm is used for calculating height field images. In a first stage a coarse matching and a dense matching is done to create overlapping range images. In a second stage those overlapping images are fused to non-overlapping height field images which result in the final DSM. Fig. 4 shows the resulting DSM of the central Gleisdorf area. Within this area six small samples were chosen to investigate the quality of the digital surface models. The positions of these samples are marked up in Fig. 4.

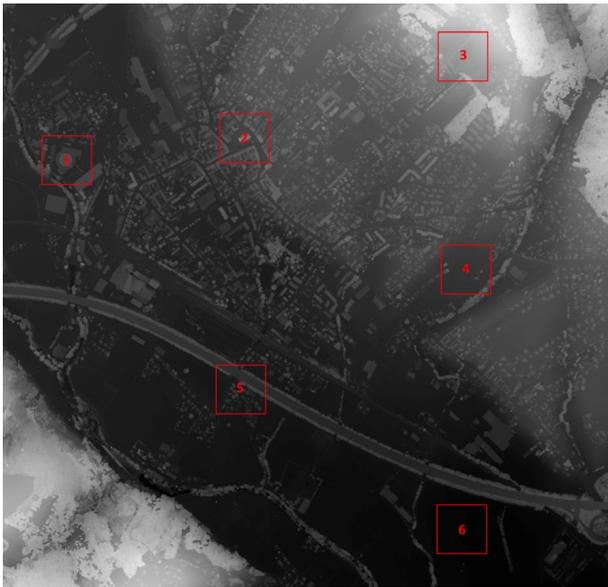


Fig. 4: DSM of the central Gleisdorf area including the six samples for detailed analysis.

2. DETAILED ANALYSIS AND RESULTS

The following analysis is performed in order to investigate the accuracy of the entire photogrammetric process and the resulting DSM data sets. The first quality measure of the process can be given from aerial triangulation. Then we compare the DSM data derived from the two sets of images taken under the different flying direction North to South vs. East to West (NS data set and EW data set).

All available image data from the flight missions show a GSD of 10 cm. The DSM datasets were processed at a mesh size of 20 cm, thus the DSM grid spacing corresponds to two pixels of the source images.

Matlab (The MathWorks, Inc.) was used to compute the specific test data and the height profiles. The overlay of checkpoints and DSM was performed in ESRI ArcGIS.

2.1 Aerotriangulation results

The aerotriangulation process includes the automatic tie point matching based on a manifold of up to 15 images for each of the two data sets, the manual measurement of ground control points in the images and the adjustment of the redundant set of observations by means of the least squares method. This process was done by using the MatchAT software product. In addition to this a camera specific set of auto calibration parameters was introduced and estimated by using the BINGO software package.

The small image distortions which were computed during the auto calibration were then introduced into the image post processing and therefore the geometric image quality and the DSM computation could be improved.

Sensor	Project		Accuracy	
	GSD (cm)	HAGL (m)	Sigma _o (µm)	Rmse z (m)
UCX	10	1400	0.80	0.035

Table 1. Overview of the main AT parameters and results. The good vertical accuracy corresponds to relative accuracy of 1/40 000.

2.2 Comparing the DSM datasets against ground truth

The initial test of the geometric quality of the DSM data was performed by comparing DSM data vs. ground truth. For this test a subset of 16 GCPs of the Gleisdorf area was selected and back projected into the DSM data sets. The vertical value of the DSM mesh near to the specific GCP was then extracted and compared against the z coordinate of the GCP. The results correspond to the expectations from the good results of the aerial triangulation. The standard deviation of the 16 height differences (13 differences in the case of data set NS) was in the range of 1 pixel (~ 10 cm).

PN	Z GCP	Z NS	zDiff NS	Z EW	zDiff EW
P 1	352.688	352.510	0.178	352.518	0.170
P 2	367.966	367.823	0.143	367.843	0.123
P 3	361.062	361.032	0.030	361.002	0.060
P 4	384.855	384.771	0.084	384.702	0.153
P 5	373.465	-	-	373.490	-0.025
P 6	359.916	359.889	0.027	359.884	0.032
P 7	349.424	349.423	0.001	349.389	0.035
P 8	348.021	348.001	0.020	347.896	0.125
P 9a	352.230	352.097	0.133	352.140	0.090
P 9b	352.594	352.656	-0.062	352.532	0.062
P 10	348.227	348.248	-0.021	348.233	-0.006
P 11	364.108	-	-	364.125	-0.017
P 12	360.394	-	-	360.509	-0.115
P 13	344.379	344.457	-0.078	344.380	-0.001
P 14	342.803	342.938	-0.135	342.956	-0.153
P 15	345.585	345.837	-0.252	345.751	-0.166
		MIN	-0.252		-0.166
		MAX	0.178		0.170
		Mean	0.005		0.023
		Stdev	±0.119		± 0.102

Table 2. Comparison of z-Values of two different DSM datasets against check points of the Gleisdorf area (all data given in meter).

Table 2 lists all GCPs with their given z-values (Z GCP) and the corresponding z-Values extracted from the two DSM datasets (Z NS and Z EW). The two additional columns show the differences between Z GCP and z-values from the DSM (z Diff NS and z Diff EW). Finally Table 2 presents minimum and maximum value, mean value and standard deviation which were computed from the set of vertical differences. The relative accuracy in z corresponds to 1/12000 for data set NS and 1/14000 for data set EW. The location of the 16 GCPs is shown in Fig. 5.

Comparing the z-values of the two DSM data sets at the 13 positions of the common GCPs we computed a mean difference of 0.035 m (1/3 of a pixel) and a standard deviation of 0.064 m. Interpreting the mean difference as a global offset we reduced this value from each height difference. The standard deviation of that set of 13 values was then only 0.053 m which is about 1/2 of a pixel of the source images.

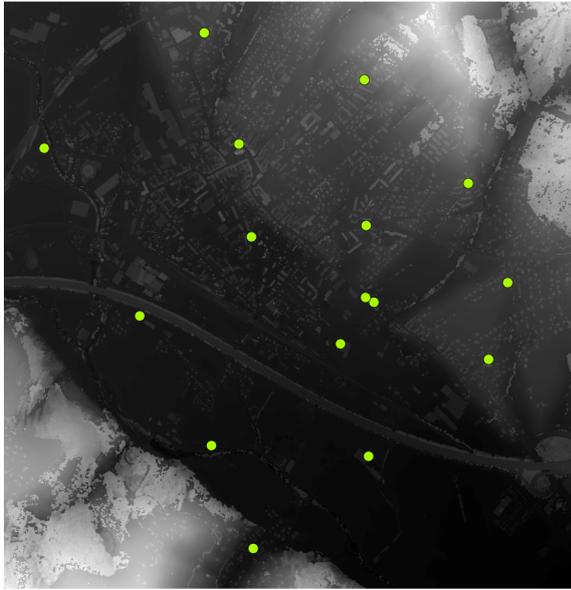


Fig. 5: Set of 16 GCPs mapped onto the DSM of the Gleisdorf area.

2.3 Comparing DSM datasets and measured z-values

For this test manually measured points were used to proof the quality of the DSM in local regions (sample areas). The measurements were done within the Match AT software package and the coordinates of the selected points were computed by means of a bundle adjustment.

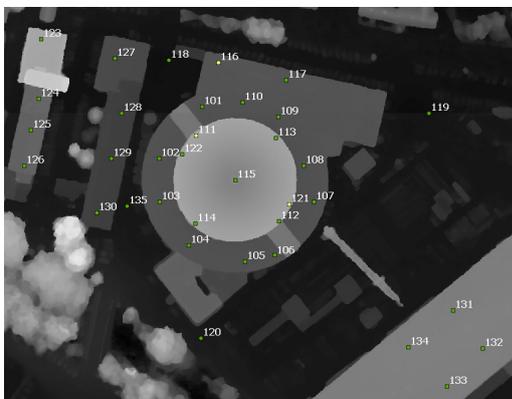


Fig. 6: Sample area 1 showing different buildings of an industrial site. A set of 35 points was measured in the two DSM datasets and z values were compared.

The result of this test is presented in Table 3. The two sets of 35 points show vertical differences of a magnitude of less than 1 pixel (StdDev of ± 0.069 m and ± 0.058 m). Differences between measured values and values extracted from the DSM data sets were used for this test. The two points 115 and 125 did show larg differences in the NS DSM data set and were rejected.

Point	Zmeasured	Z NS	DiffNS	Z OW	DiffOW
101	357,128	357,125	0,003	357,165	-0,037
102	357,190	357,159	0,031	357,174	0,016
103	357,167	357,134	0,033	357,213	-0,046
104	357,135	357,140	-0,005	357,167	-0,032
105	357,153	357,211	-0,058	357,160	-0,007
106	357,195	357,127	0,068	357,158	0,037
107	357,101	357,070	0,031	357,129	-0,028
108	356,962	356,951	0,011	356,998	-0,036
109	357,143	357,126	0,017	357,166	-0,023
110	357,102	357,057	0,045	357,147	-0,045
111	366,191	366,109	0,082	366,089	0,102
112	366,088	366,064	0,024	366,115	-0,027
113	366,038	366,096	-0,058	366,057	-0,019
114	366,047	366,067	-0,020	366,019	0,028
rej. 115	365,231	364,450	0,781	365,25	-0,019
116	357,126	357,263	-0,137	357,305	-0,179
117	357,362	357,352	0,010	357,336	0,026
118	350,969	350,859	0,110	351,011	-0,042
119	350,729	350,787	-0,058	350,749	-0,020
120	351,326	351,365	-0,039	351,291	0,035
121	366,190	366,022	0,168	366,076	0,114
122	366,114	366,055	0,059	366,153	-0,039
123	365,061	365,118	-0,057	365,148	-0,087
124	359,020	358,919	0,101	359,053	-0,033
rej. 125	359,007	359,915	-0,908	359,100	-0,093
126	359,044	359,026	0,018	359,048	-0,004
127	356,462	356,359	0,103	356,412	0,050
128	355,258	355,218	0,040	355,230	0,028
129	355,252	355,238	0,014	355,240	0,012
130	355,243	355,231	0,012	355,296	-0,053
131	362,021	361,912	0,109	361,968	0,053
132	362,059	361,957	0,102	362,118	-0,059
133	362,056	362,116	-0,060	362,003	0,053
134	362,034	361,963	0,071	362,121	-0,087
135	350,914	350,866	0,048	350,880	0,034
		MIN	-0.137		-0.179
		MAX	0.168		0.114
		Mean	0.025		-0.010
		StdDev	± 0.069		± 0.058

Table 3: Z-coordinate values of 35 points are computed from manual image measurements and compared to Z-values from DSM data sets. All units are meter. Point 115 and 125 were classified to be out layers and were rejected.

2.4 Comparing DSM datasets computed from different sets of aerial images (flying directions NS vs. EW)

In the following section we show the vertical differences of six sample areas (cf. Fig. 4) computed from the two DSM data sets. Each sample area consists of 1 000 000 height values, the size of each area is 200 m by 200 m. Each area has a specific kind of cultivation. Sample area 1 consists of industrial buildings with flat roofs, area 2 consists of residential buildings in the center of the village of Gleisdorf including the gothic style church, area 5 includes parts of a motor highway and area 3, 4 and 6 are a mixture of agricultural flat land and forest areas.

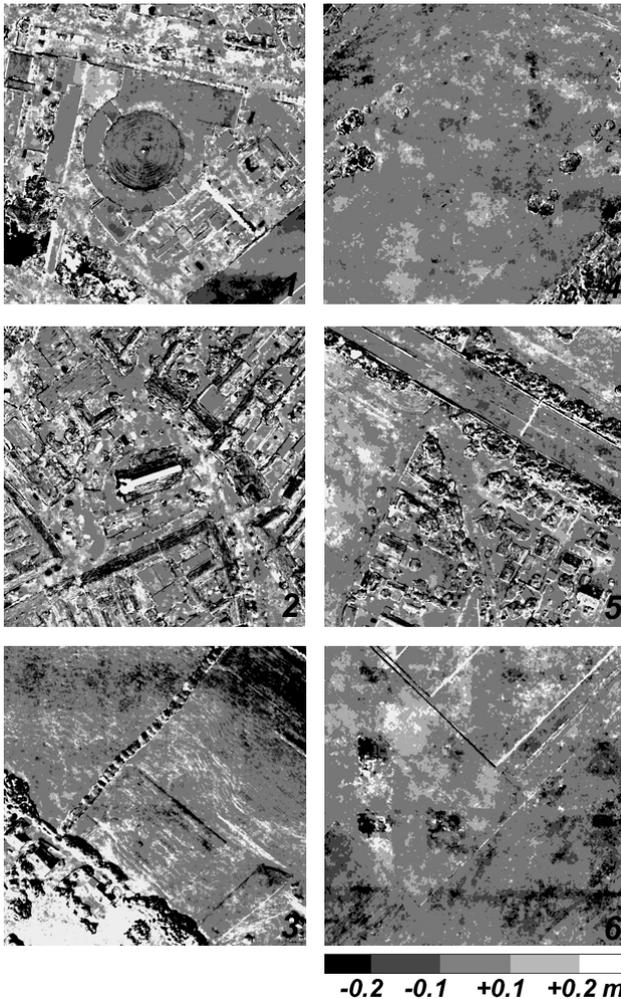


Fig. 7: Sample area 1 to 6 and the height differences computed from the two DSM data sets. Height differences are coded by different grey levels representing 5 classes of magnitude.

Sample	$ Dz < 0.3 \text{ m}$	$ Dz < 0.2 \text{ m}$	$ Dz < 0.1 \text{ m}$
1	86.4 %	74.9 %	45.8 %
2	81.6 %	70.0 %	43.4 %
3	84.2 %	74.2 %	50.0 %
4	96.1 %	93.2 %	73.1 %
5	85.1 %	76.5 %	49.9 %
6	97.7 %	91.7 %	60.9 %
1 to 6	88.5 %	80.1 %	53.9 %

Table 4: Vertical differences within each sample area are mapped into three classes. The classes represent maximum height differences of 0.3 m, 0.2 m and 0.1 m.

The comparison of the 1 000 000 height values of each of the 6 sample areas offers a measure of the noise of the dense DSM data sets. Table 4 shows, that more than 50 % of the vertical differences are smaller than 1 pixel (0.1 m) and more than 80 % are smaller than 2 pixels (0.2 m). It is also obvious, that edges of buildings, forest areas and moving objects (cars on the highway) have their specific impact on the data set. In order to overcome that issue and compute a more meaningful quality measure of the vertical differences of the two DSMs, profiles were defined and height differences along such profiles were investigated. A vertical profile through sample area 4 shows a magnitude of 0.1 m to 0.15 m and a standard deviation of $\pm 0.075 \text{ m}$. The total number of observations is 1000.

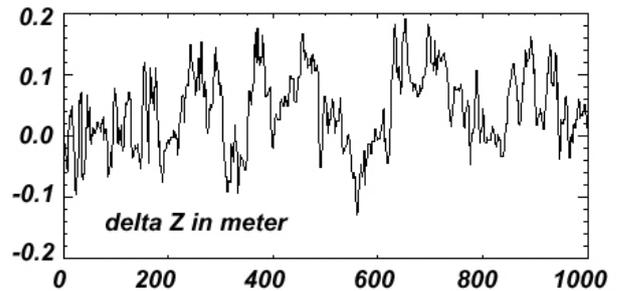


Fig. 8: Vertical profile through sample area 4. The magnitude of vertical differences of the two DSM data sets is illustrated.

3. DISCUSSION

In this contribution we use two sets of digital aerial images from a flight mission over a test area. Different flying directions (North to South vs. East to West) were planned for the aerial mission and images were taken at large overlaps. Exterior orientation data were derived through the aerial triangulation procedure. Based on the two sets of images a dense DSM at a mesh size of 20 cm (2 Pixels) was computed. The quality of the two DSM data sets was investigated by

- Comparing ground truth and DSM data
- Comparing manually measured height values and DSM Data
- Comparing the two DSM data sets.

The quality level of the aerial triangulation was quite high. A σ_o value of $\pm 0.8 \mu\text{m}$ (1/9 pixel) could be achieved. The standard deviation of z values computed from the bundle adjustment was at a 1/3 pixel level ($\pm 0.035 \text{ m}$). When comparing the vertical values of the DSMs with the ground control points we observed differences with standard deviations of $\pm 0.12 \text{ m}$ and $\pm 0.10 \text{ m}$ (~ 1 Pixel).

The vertical accuracy within a smaller sample area was then investigated by comparing manual measurements and DSM data. From a set of 35 points we computed a standard deviation of the vertical differences of $\pm 0.069 \text{ m}$ and $\pm 0.058 \text{ m}$ which is 70 % of the GCP vs. DSM test.

Comparing these results it looks like if the overall vertical accuracy of the DSM datasets is about 1/3 and the local DSM accuracy is about 1/2 of the accuracy of the highly redundant aerial triangulation

In addition to the point to point tests of the digital surface models we have compared the two DSMs through six selected sample areas and observed larger variations in the resulting quality measures. Depending on the kind of cultivation or vegetation we observed about 50% of the height differences less than 0.1 m, 80% less than 0.2 m and almost 90% less than 0.3 m. Problems raised with any kind of larger or smaller forest areas as well as buildings and other man made objects with discontinuities of the digital surface. Such discontinuities caused differences in z not as an indicator for inaccurate height measurements but more for a lack of coincidence of the horizontal positions of the discontinuity in the two DSM data sets. However, these effects have been ignored during our tests.

4. FUTURE WORK

The overall accuracy of the DSM data computed automatically from UltraCam images is very promising. Even if space for improvement is still there we see that such data have a value within several applications like digital ortho image production or 3D object reconstruction. Figure 9 gives a small insight view on how such orthophotos can look like. The resulting ortho image sample has a GSD of 10 cm and was produced on top of the DSM at 20 cm mesh size. The sample was processed completely automatically without any manual interaction.



Fig. 9: Samples of a automatically processed orthoimage of the central area of Gleisdorf.

Beyond the ortho image production the dense DSM data is the basic dataset for any three dimensional city modeling task as the Virtual Earth Initiative of Microsoft. About 100 US-cities extracted from UltraCam aerial images are already online on Virtual Earth. One example among many others is the city of Philadelphia (cf. Fig. 10).

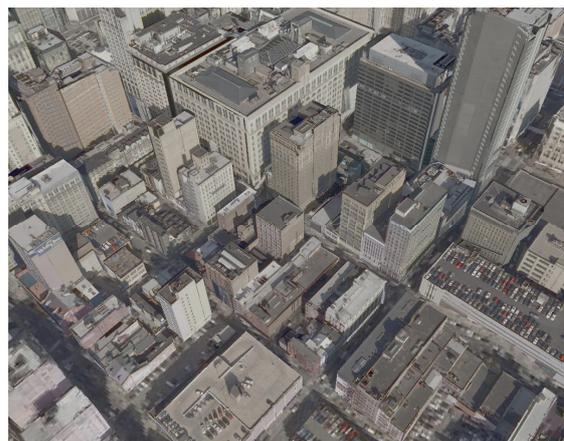


Fig. 10: 3D model of the city of Philadelphia, automatically computed from UltraCam images at high overlap.

The three dimensional digital model of the city was extracted automatically from UltraCam images, the single source for geometry and photo texture. The images were taken at high overlap, enabling automatic, redundant and robust processing.

REFERENCES

- Gruber M., Franz Leberl, Roland Perko (2003) *Paradigmenwechsel in der Photogrammetrie durch Digitale Luftbildaufnahme?* Photogrammetrie, Fernerkundung, Geoinformation, Wien 2003.
- Kremer J., Gruber M. (2004): Operation of The Ultracam Together with CCNS4/Aerocontrol – First Experiences and Results. The International Archives of Photogrammetry and Remote Sensing, VOL XXXV/1, p. 172 ff., July 2004, Istanbul, Turkey.
- Leberl F., M. Gruber, M. Ponticelli, S. Bernoegger, R. Perko (2003) *The UltraCam Large Format Aerial Digital Camera System*. Proceedings of the ASPRS Annual Convention, Anchorage USA, 5-9 May 2003.
- Zebedin, L., A. Klaus, B. Gruber-Geymayer, K. Karner (2006) *Towards 3D map generation from digital aerial images*. International Journal of Photogrammetry and Remote Sensing, Volume 60, p 413ff, 2006.

CLOSE RANGE PHOTOGRAMMETRY USED FOR THE MONITORING OF HARBOUR BREAKWATERS

M. Hennau^{a,*}, A. De Wulf^b, R. Goossens^a, D. Van Damme^a, J. De Rouck^b, P. Hanssens^c, L. Van Damme^d

^a Department of Geography, Ghent University, Krijgslaan 281 S8, 9000 Ghent, Belgium - Marc.Hennau@ugent.be

^b Department of Coastal Engineering, Ghent University, Technologiepark Zwijnaarde 904, 9052 Zwijnaarde, Belgium - Julien.DeRouck@ugent.be

^c VLG-LIN-AWZ-Waterwegen Kust, Rederskaai 50, 8380 Zeebrugge, Belgium - Paul.Hanssens@lin.vlaanderen.be

^d VO-MOW-Maritieme Toegang-Cel Kusthavens, Vrijhavenstraat 3, 8400 Oostende, Belgium - Luc.Vandamme@lin.vlaanderen.be

KEY WORDS: Digital, Reconstruction, Close Range, Photogrammetry, Measurement, Method

ABSTRACT:

Breakwaters are constructed to protect harbours against the destructive power of the sea. The top layer of rubble mound breakwaters is often composed of clear-cut concrete armour units. With the aim to maintain the integrity of these breakwaters, the units have to stay within their original pattern. Therefore, breakwaters have to be accurately monitored in order to detect any shift in position of any of the concrete armour units. A specific methodology, combining surveying and close range photogrammetric techniques, has been developed by Ghent University to perform the monitoring of the concrete armour units on breakwaters and was tested at the seaport of Zeebrugge.

1. INTRODUCTION

Worldwide, breakwaters are erected to shield outer seaport facilities from the destructive power of waves. In 1976, a rubble mound type of breakwater was selected for the protection of the expanding outer port of Zeebrugge (Fig.1).

The main constructions, which required state of the art civil engineering methods, were finished in 1985. The top layer of the breakwaters is composed of clear-cut concrete armour units which were specially positioned in order to dissipate the wave energy (Fig.2).

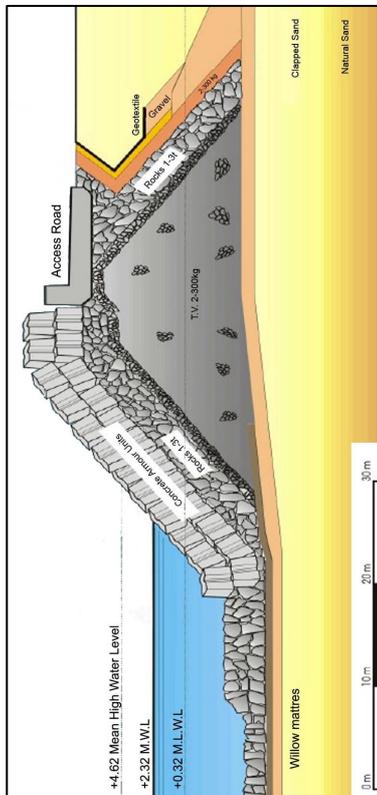


Figure 1. Cross section of the breakwater at Zeebrugge

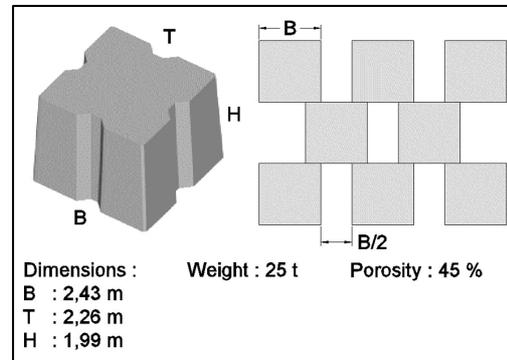


Figure 2. Concrete armour unit and front view of the original placement pattern

Unfortunately, subsidence of the different layers is unavoidable over the years. With the aim to maintain the integrity and function of the breakwaters, the pattern changes in the cover layer have to stay within acceptable boundaries and therefore the concrete armour units must be carefully monitored.

A specific methodology, combining surveying and close range photogrammetric techniques, has been developed by Ghent University to perform the monitoring of concrete armour units and was tested at the seaport of Zeebrugge, on a 500m long part of the western breakwater. During development, special attention was given to the optimization of the geometrical accuracy of the different steps. This paper will discuss the separate semi-automated processes of the developed methodology in chronological order.

* Corresponding author

2. METHODOLOGY

2.1 Survey

2.1.1 Topography: First of all, a new reference network, consisting of a chain of topographic nails, had to be materialized on the breakwater and related to the Belgian Lambert '72 planimetric coordinate system (BD72) and altimetric TAW level. A series of photogrammetric beacons, with centred topographic nails, was subsequently added to the reference network along the test site. The planimetric positions of the new reference points and the centres of the photogrammetric beacons together with the existing harbour reference poles were measured using the topographic forced centring surveying technique. Angle and distance measurements were carried out with a Leica TC1610 total station (1,5'' angle accuracy and 2mm + 2ppm distance measurement accuracy, according to DIN 18723). Both reference points and photogrammetric beacons were leveled using a Zeiss DIN11T (nominal accuracy 0,3mm/km) in combination with an invar rod. Based upon the known coordinates of the harbour reference poles, the new reference network was transformed into the Belgian coordinate system.

2.1.2 Photography: The second part of the fieldwork consisted of close range photogrammetric shots along the test site. A series of high-resolution digital photos covering the breakwater was taken from a telescopic mobile crane (SK598-AT5) with a Canon EOS-1ds fitted with a lens with a calibrated focal length of 24,513 mm. The photos were taken nearly vertically with theoretic interval distances of 25m and overlaps of 60% in order to form a photogrammetric strip with its simulated flight line centred above the top of the mound and parallel to the breakwater on an altitude of ca. 50m above the access road (Fig.3).

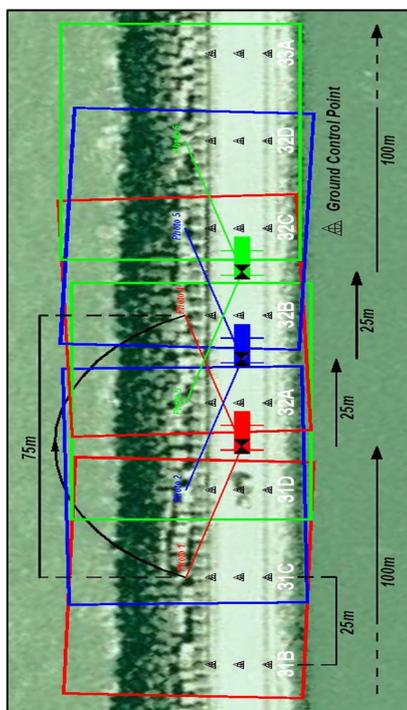


Figure 3. Photogrammetric strip

Given these parameters, a photo scale of 1:1800 and shot angles of 74 by 53 degrees were achieved. Simultaneously, approximate coordinates for the centre of projection of the camera were measured by a differential code satellite positioning system (Leica SR-20 series). The photogrammetric beacons were materialized every 25m crosswise on the access road and central wall in order to provide the overlapping photo couples with at least six common ground control points.

2.2 Digital processing

2.2.1 Photogrammetry: Digital photogrammetric processing was carried out on the photogrammetric workstation Strabox* in combination with the GIS software Orbit* and its photogrammetric extension Strabo*.

Photo coordinates were measured for the ground control points, but as the beacons were all located on the upper half of the photos, photo coordinates of extra tie points were measured on the concrete armour units of the breakwater, which were situated on the lower part of the photos. Combining the approximate BD72 coordinates of the centres of projection and the BD72 coordinates of the ground control points with the photo coordinates of both ground control points and tie points, the photo strip was digitally oriented using aerotriangulation techniques and bundle block adjustment. Consequently, the breakwater was projected into the Strabox in stereo vision, which made it possible to manually measure 3D coordinates in the stereo models.

The almost cubic shaped concrete armour units were produced with clear-cut dimensions. Because of the harsh physical conditions on the breakwater, the concrete armour units are heavily weathered. Therefore, the exact vertices of the units were almost impossible to identify, excluding a straightforward measurement of a unit's position. To overcome this problem, a stepwise method was implemented to determine the positions of the units. Firstly, the most visible side of each unit, which was the top side in 99%, was considered. Secondly, two points were measured on each of the four main edges of the selected side (Fig. 4), yielding redundant geometric information, which was used for control purposes.

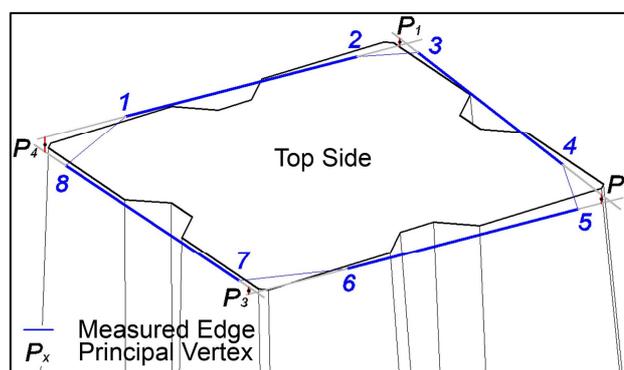


Figure 4. Top side determined by 8 points

It wasn't possible to measure eight points for every unit due to stereo occlusion. 14% of the units were determined by an alternate amount of points, ranging from seven to a minimum of three points.

* TM Eurotronics NV. Belgium

2.2.2 Automated coordinate computation: In case of eight measured points, the spatial equations of the non-intersecting edge fractions were used to compute the coordinates of four principal vertices of the considered side. In doing so, the shape of the unit's side was generalised to a square in case of a top side and to a trapezium in case of a flank side, although it must be said that the four principal vertices were non-coplanar after computation.

To compute the coordinates of one principal vertex (P₁), the common perpendicular straight line to the involved edge fractions was firstly determined (Fig. 5).

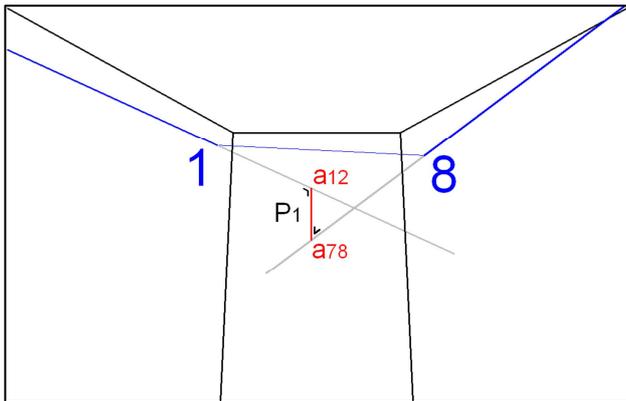


Figure 5. Constructed common perpendicular straight line

As stated in the following equation, the length of the two involved edge fractions and the distances from the points of intersection, between common the perpendicular straight line and the edge lines, to the adjacent points of the respective edge fractions were taken into account in the determination of the principal vertex.

$$P_1 = \frac{(a_{12} \times [1,2]) + (a_{78} \times [7,8])}{[1,2] + [7,8]} + \frac{(a_{12} \times [8, a_{78}]) + (a_{78} \times [1, a_{12}])}{[1, a_{12}] + [8, a_{78}]}$$

where p_1 = coordinates principal vertex
 1,2,7,8 = measured edge points
 a_{12}, a_{78} = coordinates points of intersection
 [a,b] = length straight fractions

For the edges determined by less than 2 points, some basic rules were established concerning the computation of the spatial equation of these edges and for the subsequent computation of the 2 principal vertices related to these edges. Finally, a side with theoretical dimensions was fitted on the four non-coplanar principal vertices, using a three dimensional conformal transformation.

2.2.3 Reconstructing the top layer of the breakwater, a theoretic volume model of the armour unit was fit to each computed side (Fig.6).

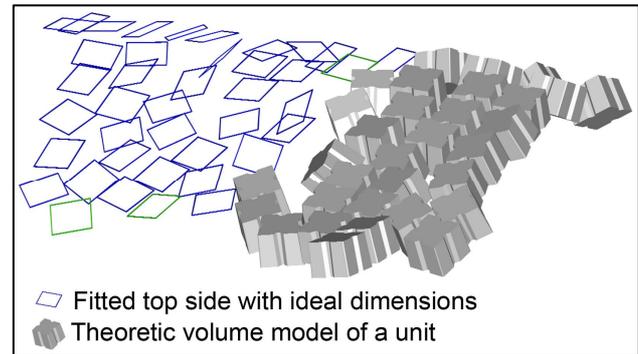


Figure 6. Reconstructed top layer

To check the accuracy of the process, a second reconstruction of a portion of the test site was performed based on an independent photogrammetric strip.

3. TEST RESULTS

3.1 Topography

After least square adjustment, the topographic survey resulted in a mean standard deviation of 9mm in planimetry for both reference points and photogrammetric beacons. The independent levelling of all new materialized points resulted in a mean standard deviation of 2mm in altimetry.

3.2 Photogrammetry

Mean standard deviations were computed for ground control points and tie points after the bundle block adjustments of the complete test strip (Table 7) and the control strip (Table 8).

Mean Standard Deviation GCPs (m)		
X	Y	Z
0,013	0,014	0,033
Mean Standard Deviation TPs (m)		
X	Y	Z
0,028	0,027	0,079

Table 7. Results bundle block adjustment test strip

Mean Standard Deviation GCPs (m)		
X	Y	Z
0,012	0,011	0,031
Mean Standard Deviation TPs (m)		
X	Y	Z
0,024	0,025	0,075

Table 8. Results bundle block adjustment control strip

3.3 Reconstruction breakwater

Residues and standard deviation for the coordinates of the four fitted principal vertices were calculated during the three dimensional conformal transformation of an ideal unit side upon each set of computed principal vertices. Coordinates of the centre of the fitted ideal side were also computed, marking each unit with a single point.

A first classification of measured units was made based on the type of measured side and on the number of measured points per side (Table 9). The altimetric position of the centre points was the basis for a second classification (Table 10). Mean standard deviations after transformation were used as comparison criteria between unit classes.

Mean Standard Deviation after Transformation (m)							
Side	#pts	#units	%	mXYZ	mX	mY	mZ
All		1301	100,0	0,096	0,041	0,037	0,074
Top	8	1118	85,9	0,094	0,040	0,035	0,074
	7	55	4,2	0,112	0,050	0,045	0,084
	6	70	5,4	0,112	0,047	0,050	0,082
	4	44	3,4	0,080	0,042	0,042	0,052
Flank	8	6	0,5	0,100	0,057	0,054	0,057
	7	1	0,1	0,081	0,042	0,039	0,057
	6	4	0,3	0,180	0,114	0,085	0,106
	4	3	0,2	0,141	0,078	0,099	0,056

Table 9.

One can notice that the relative planimetric accuracy of the measurements within the stereo models is two times better than the altimetric accuracy.

Mean Standard Deviation within Altimetric Class (m)						
Level	#units	%	mXYZ	mX	mY	mZ
0-2m	97	7,5	0,119	0,056	0,052	0,086
2-4m	204	15,7	0,126	0,055	0,050	0,096
4-6m	219	16,8	0,109	0,047	0,040	0,085
6-8m	204	15,7	0,096	0,043	0,036	0,073
8-10m	235	18,1	0,083	0,036	0,030	0,064
10-12m	342	26,3	0,072	0,027	0,028	0,057

Table 10.

Measurement accuracies are clearly lower for units at the base of the breakwater as shown in table 10.

Standard deviations after transformation are also strongly correlated to the altimetric position of the units on the breakwater as illustrated on diagram 11.

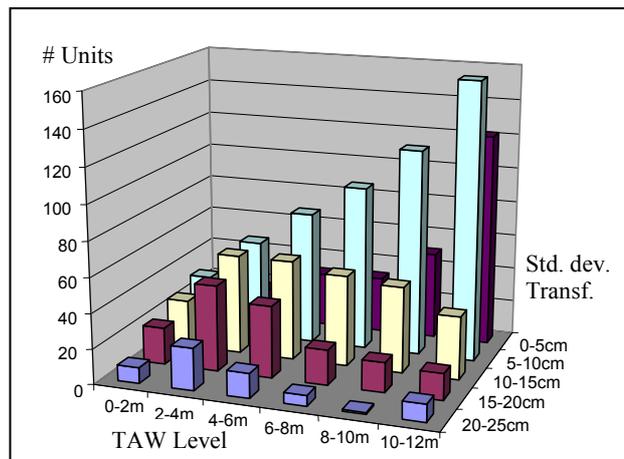


Diagram 11. Correlation between altimetric position and measurement accuracy

The results of the transformation of the independent control units and their original counterparts are fairly comparable, as shown in tables 12 and 13.

Mean Standard Deviation (m)							
Control Strip							
Side	#pts	#units	%	mXYZ	mX	mY	mZ
All		224	100,0	0,089	0,043	0,033	0,065
Top	8	203	90,6	0,090	0,042	0,033	0,067
	7	12	5,4	0,090	0,050	0,036	0,055
	6	3	1,3	0,090	0,040	0,057	0,052
	4	6	2,7	0,055	0,034	0,028	0,033

Table 12. Results control strip

Mean Standard Deviation (m)							
Original Strip							
Side	#pts	#units	%	mXYZ	mX	mY	mZ
All		224	100,0	0,097	0,044	0,037	0,074
Top	8	203	90,6	0,098	0,043	0,037	0,074
	7	12	5,4	0,086	0,047	0,038	0,060
	6	3	1,3	0,094	0,039	0,041	0,072
	4	6	2,7	0,099	0,058	0,045	0,065

Table 13. Results counterparts original strip

The control of the process was based on the computation of differences between the coordinates of the side centres of 224 units and their control counterparts (Table 14).

Mean Differences (m)				
All units				
	X	Y	Z	XYZ
	0,016	0,016	0,054	0,063
Classified according to altimetric level				
Level	X	Y	Z	XYZ
0-2m	0,082	0,062	0,077	0,139
2-4m	0,018	0,029	0,069	0,080
4-6m	0,023	0,019	0,060	0,072
6-8m	0,013	0,016	0,068	0,075
8-10m	0,012	0,008	0,041	0,046
10-12m	0,013	0,011	0,044	0,050

Table 14. Differences between original and control strip

4. CONCLUSIONS

Photo scale differences between the units on the top of the mound and at the base of the breakwater, the important height difference (+8m), in view of the photo scale, between the ground control points and the base units and the presence of seaweed at the base of the breakwater, makes the measurement of base layered units within the stereo models more difficult and less precise. Bigger standard deviations after bundle block adjustment for the tie points, measured at the base of the breakwater, and the apparent lower relative measurement accuracy in the stereo models for the lower altimetric unit classes (Table 10), support these findings.

Given the results of the control measurement, relative measurement accuracies, reflected by the standard deviation computed after fitting of an ideal unit to a measured unit, can be replicated by the applied processes. Furthermore, mean differences between original measurements and control measurements (Table 14) are well within the computed relative measurement accuracies as shown in table 9.

Given the results of the comparison between the measurements on the original strip and the control strip, it can be stated that the developed methodology can be replicated.

AUTOMATIC RECONSTRUCTION OF SHAPE EVOLUTION OF ETFE-FOILS BY CLOSE-RANGE PHOTOGRAMMETRIC IMAGE ANALYSIS

S. Hinz¹, M. Stephani², L. Schiemann³, F. Rist³

¹Remote Sensing Technology

²Photogrammetry and Remote Sensing

³Institute of Structural Design

Technische Universität München, Arcisstr. 21, 80 333 München, Germany

{Stefan.Hinz | Manfred.Stephani}@bv.tum.de; L.Schiemann@lrz.tum.de

KEY WORDS: ETFE-foils, close-range photogrammetry, shape reconstruction, bursting test, strain and stress behaviour

ABSTRACT:

The paper presents a modular photogrammetric recording and image analysis system for determining the 3D shape of Ethylen-TetraFluorEthylen-Copolymer (ETFE) foils, including their evolution over time when put under increasing air pressure. Determining the time-variable 3D shape of transparent material imposes a number of challenges – especially, the automatic point transfer between stereo images and, in temporal domain, from one image pair to the next. We developed an automatic approach that accommodates for these particular circumstances and allows to precisely reconstructing the 3D shape for each epoch as well as determining 3D translation vectors between epochs by feature tracking. Examples including numerical results and accuracy measures prove the applicability of the system.

1. INTRODUCTION

The background of this work is a civil engineering project for testing new transparent construction materials – especially Ethylen-TetraFluorEthylen-Copolymer (ETFE) foils. Membrane structures with ETFE-foils are used in building constructions in the last 25 years. Every year approximately 30 new famous ETFE-foils constructions are designed. The Allianz Arena in Munich and the Olympic Stadium Complex in Beijing for the Olympic Games 2008 are popular examples in which ETFE-foils are used for roof and facades structures. Recent projects of ETFE-foil structures show the tendency to increase the maximum span range for foil constructions. Therefore ETFE-foils are frequently stressed in the elastic range up to the yield stress or above it.

In general membrane materials are investigated with mono-axial and bi-axial tests. Bursting tests are adapted testing methods to analyze the material properties of ETFE in the viscoelastic and viscoplastic range up to the breaking point. They allow inspecting stress and strain behaviour of ETFE-foils with consideration of the rotationally symmetric deformation condition. This is one of the principal cases for pneumatic foil constructions (Moritz et al., 2005; Moritz, 2007). To test different material characteristics like stress and strain behaviour a low-cost measurement tool was set up that allows to produce densely sampled points with high accuracy in time and space domain. To this end, a photogrammetric close-range stereo system with specific image analysis software is an ideal tool. We outline the concept and realization of a modular photogrammetric recording and image analysis system for determining the 3D shape of ETFE-foils, including their evolution over time when put under increasing air pressure.

2. TEST SET-UP

In photogrammetric applications, textured and opaque materials need to be assumed in general. Both prerequisites are not met in our case. Hence, determining the time-variable 3D shape of transparent material imposes a number of challenges – most prominently, concerning the automatic point transfer from one stereo image pair to the next. To enable automatic matching of

conjugate points, a black grid was printed onto each foil. Ideally, this grid would reflect some unique code to avoid ambiguous point matches caused by the grid's regularity. For this, however, specific printers or expensive projection hardware would have been necessary, which did not fit into the project's financial budget. Instead, a sophisticated matching strategy is incorporated, which is in particular able to cope with possible matching ambiguities.

Circular ETFE foil samples with a diameter of approx. 53 cm are clamped by a locking ring on a base plate. The foils are marked by grid lines. The cushions are inflated by means of electronically controlled pressure valves, while a digital pressure sensor measures the internal pressure in the cushion. Above the foil two digital high end user cameras are positioned. They record synchronously and automatically every five seconds the deformation of the ETFE-foil (see Fig. 1). Ultimately, the 3D position of grid crosses are determined by photogrammetric means and tracked over several stereo pairs (epochs). Hence, 3D translation vectors between each pair of epochs can be calculated.

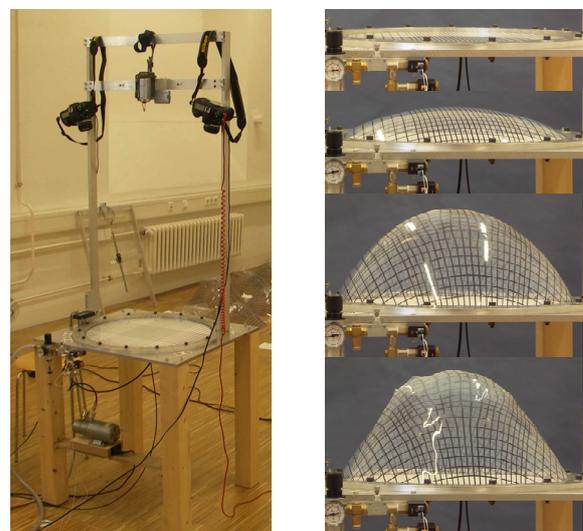


Fig.1: Experimental set-up and bursting test

3. IMAGE ANALYSIS SYSTEM

The overall processing scheme consists of following steps:

3.1 Camera calibration:

The two digital cameras of type Nikon D200 (3872 x 2592 Pixels of 6.096 micron size, corresponding to 23.6 mm x 15.8 mm image size) used are equipped with off-the-shelf 24 mm lenses. In order to exploit their full accuracy potential, calibration of the cameras in a similar environment as in our set up is a prerequisite. The parameters of inner orientation (incl. distortion parameters) of the two cameras were determined offline using a 3D test field with known ground control points. With each of the cameras 8 images of the test field were used to derive the calibrated focal length, the coordinates of the principal point and two parameters describing the distortion of the lenses. In order to obtain precise and reliable interior orientation parameters sub-pixel image measurements are desirable and the redundancy of the least squares adjustment should be as high as possible. Main results of the calibration procedure are presented in Table 1.

Parameter	Camera 1	Camera 2
Focal length [mm]	24.590	24.610
Principal point x [mm]	+0.063	-0.010
Principal point y [mm]	+0.012	+0.091
Distortion coeff. A1 [m^{-2}]	-154.9295	-157.7906
Distortion coeff. A2 [m^{-4}]	+245041.5	+257126.5
Sigma naught [micron]	1.7	1.8
Redundancy	464	466

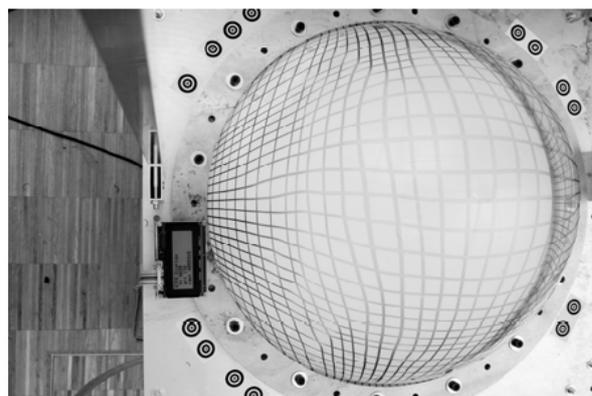
Tab.1. Camera calibration results

3.2 Relative and absolute camera orientation

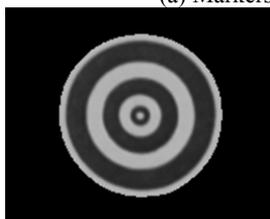
Various circular markers with known “real-world” co-ordinates have been placed in the vicinity of the ring (see Fig. 2a). These markers are automatically detected in and matched over the images. To this end, we adapted the shape-based matching algorithm proposed by (Steger, 2001) and (Ulrich, 2003).

A gradient filter is applied to the template image (Fig. 2b) and the gradient directions of each pixel are determined. The same is done for the search images. For determining the best matches, the template image is shifted over the search image and, for each position, the gradient directions of template and search image are compared. A similarity measure is calculated defined as the average vector product of the gradient directions of the template and the search image. This similarity measure is invariant against noise and illumination changes but not against rotations and scale. Hence the search must be extended to a predefined range of rotations and scales, which can be easily derived from approximate values given by the test set-up. Since we know the number N of markers, we can simply restrict the search to the N best matches. The matches are further refined to sub-pixel accuracy by fitting a second-order polynomial to the found matching parameters in a small neighborhood in parameter space.

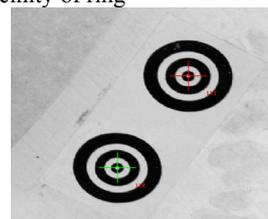
Figure 2c) shows the final accuracy of the derived center coordinates and Fig. 3 illustrates the shape-based matching procedure. The final orientation of both cameras is then determined through bundle adjustment (see (Mikhail et al., 2001) for details).



(a) Markers in vicinity of ring



(b) Template



(c) Derived center coordinates

Fig.2. Automatic detection of markers

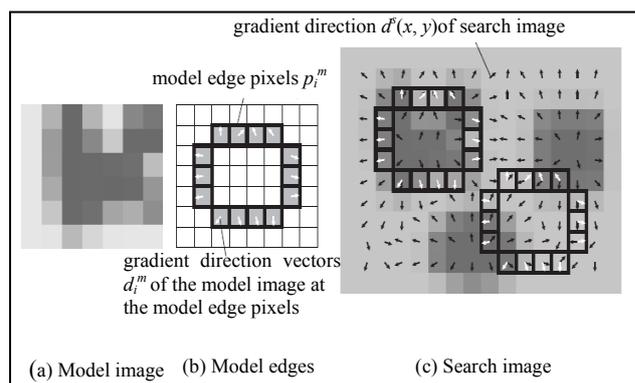


Fig. 3. Principle of the shape-based matching method, taken from (Ulrich, 2003), p. 70

3.3 Initial extraction and matching of features

In the initial (flat) state of the foil, more constraints can be incorporated into feature extraction and feature matching. Hence, this step is separated from the others. It consists of:

3.2.1 Point extraction: The crossings of the grid are determined by extracting dark lines using the approach of Steger (1998) and intersecting the endpoints and junction points of these lines. The positions of the resulting intersections are further refined by fitting a second order 2D polynomial to the grey values yielding eventually sub-pixel precision (see Fig.4).

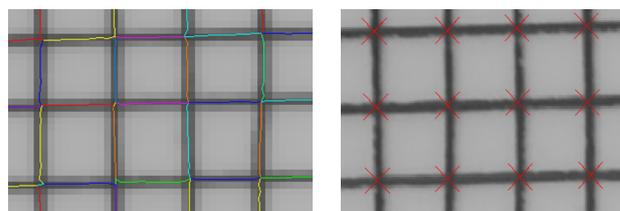


Fig. 4. Line extraction (left) and refined junction points (right)

3.2.2 Initial point matching: In the next step, conjugate feature points are determined in both images. Since the foil is flat in its initial state, a homography provides a sufficient mathematical model for point transfer and prevents a wrong solution due to over-parameterization and potential ambiguous matches. Furthermore, the search space for conjugate features can be significantly reduced in image space, since the feature points on the foil are almost co-planar with the already reconstructed 3D points of the markers. The point transfer is then calculated by estimating the parameters of the homography while maximizing the number of point correspondences using a RANSAC search algorithm, thereby the initial parameter values being derived from the camera orientation and test set-up. The optimization measure is computed from the number of points (“RANSAC inliers”), the accuracy of the estimated homography parameters and the grayvalue similarity of a small patch around the feature points. After having found a reasonably good solution through RANSAC, triangulating the resulting conjugate points yields their final 3D position in object space. Figure 5 depicts a detailed view on the resulting conjugate points.

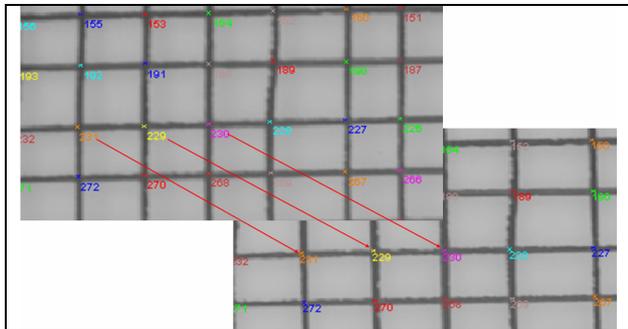


Fig. 5. Result of point transfer based on a homography

3.4 Iterative point tracking

It seems straightforward to apply the above procedure to each image pair of the sequence. However, once the foil expands, a larger search space and a less stringent mathematical model for point transfer needs to be incorporated, which – in turn – may result in more ambiguous matches. In addition, the deformation of the foil leads to a significant dissimilarity of the grayvalue patches of conjugate points when seen from the two camera position. To circumvent these problems, an alternative procedure has been implemented:

The conjugate points of the first epoch ($t = 0$) are tracked over the next epochs ($t = 1 \dots n$) separately for the images of each camera (see Fig. 6 for illustration). To this end, a small patch around each point is selected and – similarly to pyramid tracking – matched to a search area of the image of the next epoch. Here, the image patch is updated and matched to the following epoch. Updating the image patch is necessary to accommodate the increasing deformation of the grid from epoch to epoch, see e.g. Fig. 7 that shows the result of point tracking in a later epoch. Tracking a particular point aborts when no reliable match can be established.

Please notice that, in this step, it is not necessary to search for corresponding points in the left and right image. The correspondences have been established in the initial matching step and are simply kept during point tracking via point indexing.

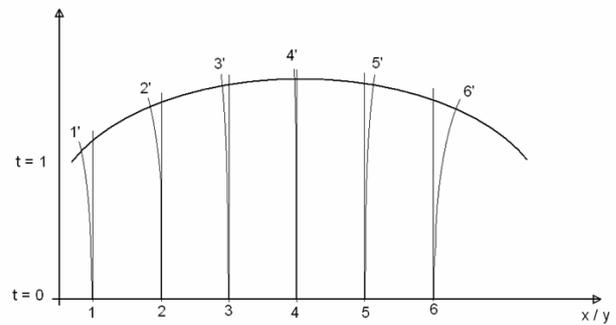


Fig. 6. Illustration of point tracking algorithm

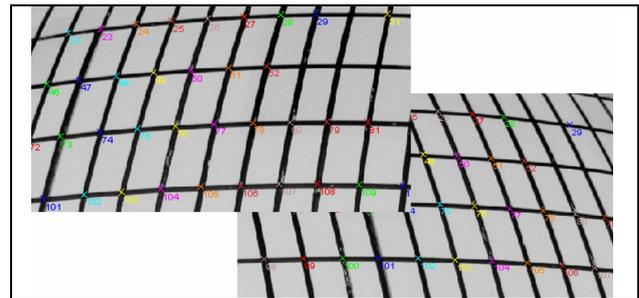


Fig. 7. Matched points in later epoch

3.5 Bundle adjustment

Bundle adjustment is considered the most appropriate tool to calculate 3D-object coordinates from multiple image coordinate measurements (Mikhail et al., 2001). The image coordinates and control point coordinates are introduced as observations into the least square adjustment process, while the exterior orientation (position and attitude) of the images and the object coordinates of homologues points are the unknowns to estimate. The interior orientation of the cameras is considered as given from the calibration process.

For relating all relevant parameters, the following well-known relation applies:

$$\begin{bmatrix} x'-x_0'-dx' \\ y'-y_0'-dy' \\ -c \end{bmatrix} = \frac{1}{m} \begin{bmatrix} r_{11} & r_{21} & r_{31} \\ r_{12} & r_{22} & r_{32} \\ r_{13} & r_{23} & r_{33} \end{bmatrix} \cdot \begin{bmatrix} X-X_0 \\ Y-Y_0 \\ Z-Z_0 \end{bmatrix}$$

where:

- x', y' measured image coordinates
- x_0', y_0' principal point coordinates
- c focal length
- dx', dy' distortion corrections
- m scale factor
- r_{ik} rotation matrix elements
- X, Y, Z object coordinates
- X_0, Y_0, Z_0 projection centre

The scale factor can be eliminated through dividing the first and second line of the equation above by the third line. Then, one obtains two highly nonlinear observation equations for each image point:

$$x' + v_x = x_0' - c \cdot \frac{r_{11}(X-X_0) + r_{21}(Y-Y_0) + r_{31}(Z-Z_0)}{r_{13}(X-X_0) + r_{23}(Y-Y_0) + r_{33}(Z-Z_0)} + dx'$$

$$y' + v_y = y_0' - c \cdot \frac{r_{11}(X-X_0) + r_{21}(Y-Y_0) + r_{31}(Z-Z_0)}{r_{13}(X-X_0) + r_{23}(Y-Y_0) + r_{33}(Z-Z_0)} + dy'$$

where:

v_x, v_y , corrections to the image coordinates x', y'
 dx', dy' radial distortion corrections

The radial distortion dr'

$$dr' = A_1 (r^3 - r \cdot r_0^2) + A_2 (r^5 - r \cdot r_0^4)$$

is split-up in two components dx' and dy' (first and second term), where r is the distance to the centre (radius), r_0 the reference radius and A_1 and A_2 the distortion coefficients obtained from the calibration procedure.

Thus, for each coordinate triple X, Y, Z of an object point to be calculated, four observations (image coordinates, two in each image) are available. As all coordinates of the markers are considered as error-free ground control information, the resulting redundancy of the bundle adjustment increases with increasing number of ground control points and object points to be determined. In order to keep the number of iterations low, suitable initial values for all unknowns are calculated at the beginning of the least squares adjustment process. Typically 650 to 880 object points are calculated iteratively by an optimized intersection of spatial rays (extended bundle adjustment) from each image pair (epoch).

The final 3D co-ordinates of each epoch's points including their accuracy and point IDs are stored to allow the computation of time-dependent 3D translation vectors between points of various epochs. The result of 3D point determination of an intermediate epoch is shown in Fig 8.

3.6 Surface Generation

The deformations of the foils are tracked through the adjusted 3D coordinates for each epoch of the points marked on the surface of the foils. Hence the behaviour of the foil can be studied and deformation parameters (strain, curvature etc) derived. On the other side these more or less arbitrary distributed 3D points for each epoch can be used to calculate a regularly grided surface using the Golden Software Package "Surfer 8". Afterwards different visualizations of the shape of the inflated foils are feasible.

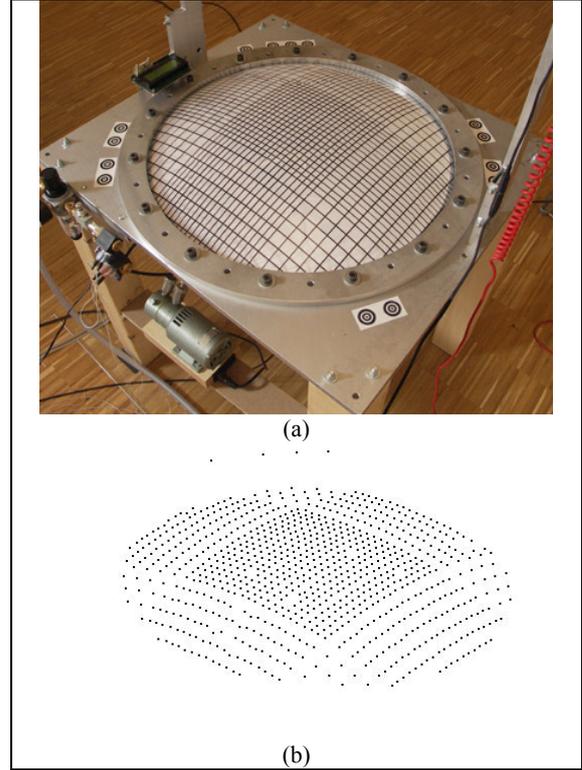


Fig. 8. 3D point determination in intermediate epoch:
 (a) current state of foil, (b) reconstructed 3D points.

4. RESULTS

With the coordinates of the matching points three-dimensional scatterplots of the deformed geometry of the ETFE cushions are determined (cf. Figs. 9, 10). They show the deformed geometry of the foil at each epoch due to the internal pressure. For the epoch shown in Fig. 10, the maximum vertical deformation of the circular ETFE-foil samples is 27,425 cm. At the beginning of the test the scatterplot exists of approx. 1000 points, while the finite element net exists of approx. 850 4-nodes elements. The curvature of the deformed geometry

$$\kappa = \frac{f''(x)}{(1 + f'(x)^2)^{3/2}}$$

is calculated from the deformation function $f(x)$ of the foils at each epoch.

To observe and describe how the foils deform during the inflating process differential geometry and continuum mechanics basics are applied to the points tracked over epochs. The Green-Lagrange-strain tensor

$$E = \frac{1}{2} (g_{ij} - G^{ij}) G^i \otimes G^j$$

describes the mechanical behaviour of the foils in consideration of the nonlinear strain measures and the large deformation. E is calculated with the difference of the metric coefficients. The covariant metric tensor g_{ij} from the local covariant basis describes the metric of the surface for the reference geometry at time t_0 . The metric of the surface at each epoch at time t_i is described with the contravariant metric tensor G^{ij} . Figures 11a-d) show a typical example after a 195 seconds inflation period and an internal pressure of 30900 Pa. The strains are depicted in addition to the deformed foil, the scatterplot and the shaded relief map.

5. CONCLUSION

The results from the photogrammetric image analysis show that, despite of the challenging environment, an average point accuracy of 1/3 pixel at image scale could be reached. In addition, the point density (starting with 100%) remains high for approx. 3/4 of the tracked image pairs and drops down, due to increasing occlusions, to the end of tracking, only.

It must be pointed out that for the accurate calculation of the nonlinear strain measures a high point accuracy of the determination of the grid point coordinates is necessary. In our case the large deformation and the transparent material complicate the determination of the points. The achieved point accuracy of 1/3 pixel at image scale enable an accurate determination of the deformations and of the curvatures. The calculated results are adequate to investigate qualitatively the strain behaviour of the foils. However, the point accuracy must be improved to calculate accurately the strains. Possible improvements are smoothing the surface, fitting the point coordinates, using a third camera or increasing the inclination of the cameras.

The developed bursting test including the photogrammetric recording and image analysis system is a possibility to investigate transparent materials, like ETFE-foil, in the viscoelastic range up to the breaking point. In addition to the determination of the deformation and curvature, the tests analyse the breaking behaviour of circular ETFE-foil cushions.

REFERENCES

Mikhail, E., Bethel, J., McGlone, C., 2001. Introduction to Modern Photogrammetry. John Wiley & Sons.
 Moritz, K., Brengelmann, T., Schiemann, L., 2005. Steel Service Station Roofs with Plastic Membranes, DETAIL 7/8.
 Moritz, K., 2007. Bauweise der ETFE-Foliensysteme, Stahlbau 76 (5)
 Steger, C., 1998. An unbiased detector of curvilinear structures, *IEEE Trans. Pattern Anal. Machine Intell.*, 20(2), pp. 549-556.
 Steger, C. (2001): Similarity measures for occlusion, clutter, and illumination invariant object recognition. In: B. Radig and S. Florczyk (eds.) *Pattern Recognition*, DAGM 2001, LNCS 2191, Springer Verlag, 148-154.
 Ulrich, M., 2003. Hierarchical Real-Time Recognition of Compound Objects in Images. Dissertation, German Geodetic Commission (DGK), Vol. C.

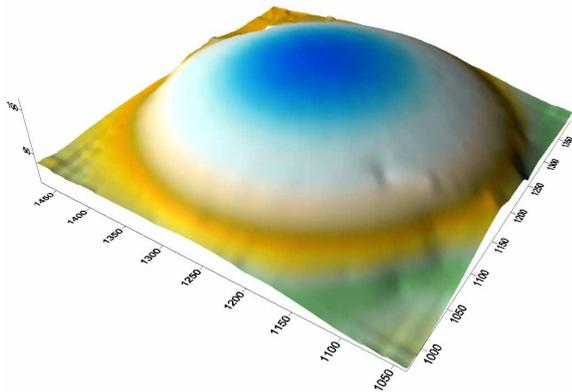


Fig. 9. Visualization of interpolated surface

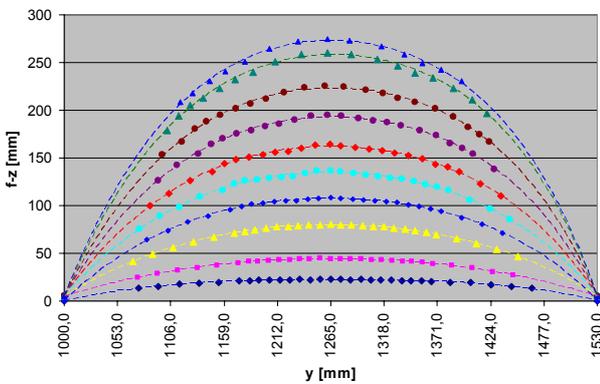


Fig 10. Deformation of the cushion for different pressures

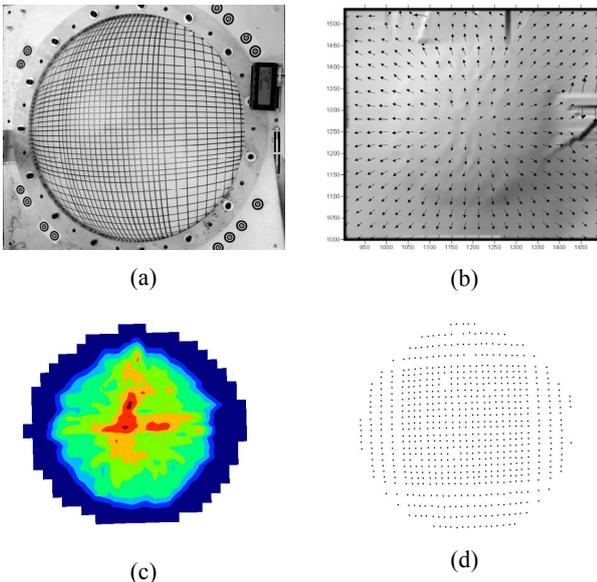


Fig. 11: Results, time t=195s, p₁ = 30900Pa, (a) deformed foil, (b) shaded relief map, (c) strains and (d) scatterplot

AUTOMATED GENERATION OF BUILDING TEXTURES FROM INFRARED IMAGE SEQUENCES

Ludwig Hoegner, Uwe Stilla

Institute of Photogrammetry and Cartography, Technische Universitaet Muenchen, Munich, Germany
Ludwig.Hoegner@bv.tum.de; Stilla@tum.de

KEY WORDS: Image Analysis, Image Sequences, Texture Extraction, 3d Building Model, Thermal Infrared, Terrestrial, Close Range, Urban

ABSTRACT:

In this paper the application of computer graphics and computer vision for texture extraction from infrared image sequences is described. These techniques normally are used in computer graphics to project a virtual scene onto the image plane. For the matching between the images and the given 3D model a strategy is presented based on the estimation of planes in the image sequence using homography. Textures are extracted using an algorithm based on the principles of ray casting to generate partial textures for every visible surface in every single image of the sequence. The textures generated from different images of the sequence belonging to the same façade are combined. By combining the intersection points of several images generated during the ray casting, a dense structure of points can be used to texture and analyse even big building complexes.

1 INTRODUCTION

The paper intends to investigate strategies for a detailed texture mapping of building façades which allow a thermal inspection. Typically, thermal inspections of building façades are carried out in single images from the observed objects. Larger building parts require several images to be analysed. An integral way of viewing buildings recorded from different images is difficult without combining those images. This problem is getting worse, when images from different cameras or views need to be combined and stored for further processing without any geometric reference.

The discussion of global warming and climate has focused on thermal inspection of single buildings on the one hand and urban environment on the other hand. Ground cameras are recording the irradiation of building facades (Klingert, 2006), for the specification of its thermal behavior. Satellite images are used for fire detection (Siegert, 2004), vegetation monitoring (Quattrochi, 1999) or the analysis of urban heat islands (Lo, 2003). Airborne IR-systems are applied for vehicle detection (Hinz, 2006, Stilla, 2002) or exploration of leakages in district heating systems (Koskeleinen, 1992). Typically, the analysis of the urban environment is done on the ground but without 3d buildings. A relatively new approach is developed by Janet Nichol (Nichol, 2005). For urban environmental quality studies, satellite IR data were combined with 3d city models. In IR images are projected onto the terrain and the building surfaces involved in the energy exchange were matched with the IR images, including vertical walls and the horizontal surfaces of the buildings seen from the satellite. This textured model permit 3d visualization for a better understanding of the factors controlling urban environment. But there is no detailed analysis of the single buildings.

In difference from areal and satellite images, ground images normally do not contain a complete building in a single image. Therefore, it is necessary to combine several images to extract the complete texture for a façade. This combination needs the knowledge of the parameters of the camera used for the record to correctly project the images into the scene. The estimation of exterior orientation from a single image works with at least 3 correspondences (3-point algorithm) between image and model (Haralick et al, 1994). Techniques for 4- and 5-point estimation are elicited by Quan (Quan and Lan, 1999) and Triggs (Triggs, 1999). For 6 and more correspondence points the Direct Linear Transformation (DLT) can be applied (Triggs, 1999). For homogeneous façade structures that approximately form a plane, homography can be adopted to detect planes in image pairs and the relative exterior orientation of the camera in relation to these planes (Hartley and Zisserman, 2000). Due to the small field of view, the low spatial resolution of the IR images and the low level of detail of the given building model, only few point correspondences between IR image and 3D model can be identified. So this paper is focused on the homography based surface estimation. The extraction of surface textures from IR images is done by a reversed variant of ray casting. Instead of calculating the pixel color from rays that collect color values in the 3d scene, the rays project the known pixel values on the intersection points of the rays and surfaces in the 3d model. The 3d coordinates of the intersection points can be directly used to generate texture coordinates. These texture coordinates are used to generate 2d textures for the surfaces. Textures generated from different images of an image sequence can be combined to achieve a higher resolution and completeness of the façade texture.

2 DATA ACQUISITION

As the wavelength of infrared light is much longer than the visible spectrum, other optics and sensors are necessary to record infrared radiance. Current IR cameras cannot reach the optical resolution of video cameras or even digital cameras. The cameras that were used for the acquisition of the test sequences offer an optical resolution of 320x240 (FLIR SC3000) and 320x256 (MerlinMID) pixel with a field of view (FOV) of only 20° (Fig. 1). The SC3000 is recording in the thermal infrared (8 - 12 μm), whereas the MerlinMID is a midwave infrared camera (3 - 5 μm). For the methods used in this paper, both cameras were mounted on a rotatable and shiftable platform on the top of a van. Images in the mid-wave infrared are directly affected by the sunlight as in addition to the surface radiation caused by the building's temperature the radiation of the sun is reflected. In long-wave infrared the sun's influence appears only indirect, as the sun is not sending in the long wave spectrum, but of course is affecting the surface temperature of the building.



Figure 1: Used camera system: midwave (3-5 μm) and the longwave (8-12 μm) infrared camera (see the left and middle camera) and video camera right

To inspect the temperature loss of a building, the environment should be cold, where as the heating of the building should be running to observe a relevant temperature flux. Optimal time is after sunset and before dawn in spring. To minimize occlusion caused by vegetation a date before foliage is to be preferred. Caused by the small field of view and the low optical resolution it is necessary to record the scene in oblique view to be able to record the complete facades of the building from the floor to the roof and an acceptable texture resolution. That is why the optical resolution of the images is not constant. Image parts that show the complete building have a lower resolution than image parts that show only the first floor of the building. When several images are combined, the optical texture quality of the final surface texture is high in the first floor and decreases to the top of the façade. Due to this, the image sequences are recorded with a frequency of 50 frames per second. To minimize holes in the textures due to occlusion caused by the oblique view, every façade is recorded with a view forward looking and a view backward looking. The viewing angle related to the along track axis of the van must be constant. The position of the cameras is to be recorded with GPS.

3 TRANSFORMATION OF 2D IMAGE DATA TO 3D POINTS AND TEXTURE COORDINATE

3.1 Position estimation and image matching

The camera parameters must be known in order to project the 3D model onto the image plane. The interior orientation of the camera is determined by a calibration. The position of the camera is recorded during the image acquisition. Pan, tilt and roll angles of the camera are estimated from ground control points given by the vertices of the 3D model. Caused by the high buildings the recorded GPS signal is inadequate and only allows a position accuracy of about 5 meters. This is to inaccurate to use the position directly for the projection and does not give the pan, tilt and roll angles.

Instead of calculating the position for individual key frames, the image sequence is used to orient the images. For homogeneous façade structures that approximately form a plane, homography can be adopted to detect planes in image pairs and the relative exterior orientation of the camera in relation to these planes (Hartley and Zisserman, 2000). Many buildings have an approximately planar façade. Assuming that façade structures lie in a plane, a correspondence between points of two subsequent images can be found using a homography matrix H . Points of interest are found searching eigenvalues and performing a non-maxima-suppression. The homography matrix is then calculated for corresponding points selected by RANSAC (Fischler and Bolles, 1981). From subsequent image pairs with frame distance d a set of homographies are calculated to allow the estimation of a trajectory of the camera relative to the façade plane which is averaged by combining the planes of the homographies. For the first image pair, the initial camera position is used to estimate a façade plane that is close to the corresponding 3d model surface. This strategy works well for homogeneous images with only one façade covering most of the image.

3.2 Selecting the visible objects

Ray casting searches for intersection points between the rays sent through the pixels of the image plane and all surfaces of the 3d scene. Using a typical resolution for an IR image of a video sequence, there are 320 x 240 pixels per image (values of camera FLIR SC3000 used for the data acquisition). This means at least 76800 pixels which have to be checked against all surfaces. The surfaces of the 3d models consist of triangles. A building façade consists of at least two triangles to form a rectangle. In result, for every surface there are at least 153600 intersection points to calculate. When using several rays per pixel to achieve a higher accuracy and avoid holes in the resulting textures, this number gets even higher. It seems reasonable to minimize the number of surfaces, which have to be checked for intersections. To achieve this, frustum culling is used for all polygons of the surfaces of the given 3d model using the camera parameters estimated in the homography. This way, only surfaces being partially within the field-of-view are processed in the projection and texture extraction steps. A second way to minimize the number of surfaces for ray casting is the back face culling. That means, that surfaces within the field of view but invisible can be removed from processing. Back faces are the back sides of buildings completely occluded by the building's front side. In computer graphics the distinction between front and back faces is made by the order of the vertices defining a triangle. Unfortunately, the models that are used for the test in our project, have not

always correctly defined triangles to distinguish between front and back faces. That is why back face culling is not used in this project, but of course can be used for 3d models with correct back face definition.

3.3 Projection of image pixels into the 3d scene

Depending on aspect and position of the camera view, parts of the buildings are invisible due to self occlusion or occlusion from other objects. For every pixel of the image of the virtual camera the corresponding surface point is searched. First, for every surface a plane equation is calculated to receive the depth value and texture coordinates. This plane is used for a ray casting (Foley 1995), where every pixel is projected into the scene and assigned to an intersection point of the plane which has the smallest depth value along the ray from the camera through the pixel. For the intersection point of this plane texture, the texture coordinates are interpolated from the texture coordinates of the vertices of the plane. The texture coordinates count from $(u,v) = (0,0)$ at the left lower vertex of the surface and are going up to $(u,v) = (1,1)$ at the upper right vertex. The ID of the intersected surface and the texture coordinates of the intersection point are returned to the pixel of the infrared image. After this 3D to 2D transformation of the model surfaces into the image plane of the IR image a 2D transformation is carried out to transform the IR image pixels to texture coordinates of the model surfaces. After this process the pixel values of the IR image have been transformed to texture coordinates of points on the model surfaces.

A further 2D transformation is conducted to transfer the texture points to pixel coordinates for the surface textures. At first, the individual pixel coordinates of the texture are transferred into texture coordinates of the surface. Then, their interpolated values are calculated by means of a bilinear interpolation. If a pixel has only three surrounding points in u and v direction, the pixel value is interpolated using barycentric coordinates. If the pixel is outside the triangle defined by the three surrounding points, it is outside the visible part of the façade. Pixels with only two or less surrounding texture points also are outside the visible part of the façade and left in black.

3.4 Combination of the partial textures of one surface

For composing a complete texture for a building façade, it is necessary to combine several partial textures generated as described in 3.1 to 3.3. Because the camera parameters are containing a small error even after correction, like mentioned in 3.1, the surface textures generated from different images of the same sequence may not be congruent. But, as the input image sequence is recorded with several frames per second, in our case study with 50 frames per second, and the viewing angle related to the along track axis of the van is constant, the disparity of two adjacent images is very small in movement as well as in rotation and so is the possible mismatch between the extracted surface textures of those two IR images.

Due to this, corresponding points in both surface textures are searched within a small area of only several pixels for every point. After two surface textures are matched, the intensity values for the combined texture have to be calculated. For an image sequence, the resolution of the surface textures is not constant. The resolution decreases in the viewing direction of the camera caused by the perspective view. Texture points, which have a bigger distance to the camera, have a lower spatial resolution. When recording forward, the visible part of the façade texture of each following image has a higher spatial

resolution than the image before, but does not show the complete part of the image before. Figure 2 illustrates the resolution distribution of a combined surface texture.

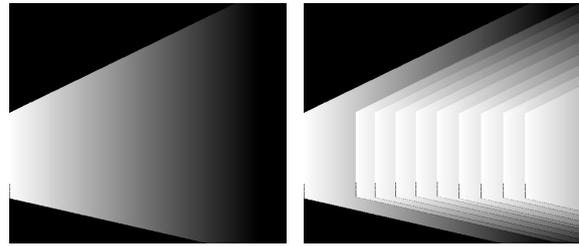


Figure 2: left: Resolution image of a single partial texture, right: Resolution image of a combination of several partial textures

For backwards view, every previous image as a higher spatial resolution, but does not show the complete part of the following image. This fact allows the use of a very simple texture combination method. For forward viewing, initially the first partial surface texture is copied to the combined texture. The second partial surface texture is then copied to the combined texture and overrides the pixel copied from the first texture, that are also present in the second one. But, as mentioned before, all pixels of the second texture have a higher resolution than the first texture. For the first texture, only the pixels that are not part of the second texture remain in the combined texture. This procedure is continued for all partial surface textures of the image sequence. In the end, the combined final texture has the highest possible resolution that can be achieved from the input image sequence. For the backward viewing image sequence, the last partial surface texture is initially copied to the combined texture and then the partial surface textures of the sequence are added in reverse order.

Generating the final texture in this way leads to a texture with the highest resolution for every texel that is extractable from all partial textures. Figure 3 shows a continuous distribution of the resolution of the complete texture of the façade assuming a high image rate and a low velocity.

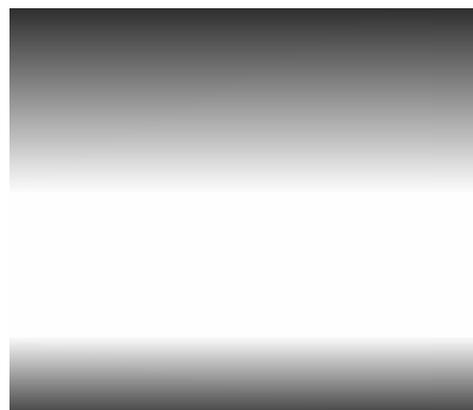


Figure 3: Resolution image of the final texture

The white region in the middle corresponds to the first floor of the building, which lies in the view axis of the camera and can be seen at the left high resolution edge of all partial textures. Due to the oblique view the resolution decreases downwards and upwards. These texture parts cannot be seen at the left edge

of the images, but more to the right, where the resolution decreases. This effect depends on the angles of the camera view towards the camera path and towards the façades. The bigger the angle between the viewing direction of the camera and the surface normal of the façade, the bigger gets this effect.

Because the appearance of IR signatures is changing over time and depends for example on the weather conditions, the combination of textures from different records at different days or daytimes to create a new combined texture is very difficult. That is why normally only textures of one record session should be used to create new textures.

The final combined textures of all visible façades are assigned to building surfaces and stored together with the 3d model in a hierarchical object oriented database for further processing like feature extraction. The 3d intersection points are also stored as they can be used later to create geometry if the feature detection in the textures identifies geometry.

4 PROCESSING OF TEST DATA

For the experiments with the camera system mentioned in section 2 a building complex is chosen that shows long façades as well as discontinuous façade structures in narrow streets. The given 3d building model containing the surface polygons of the façades (Fig. 4) is combined with the longwave infrared image sequence of the SC3000 camera (Fig. 5)

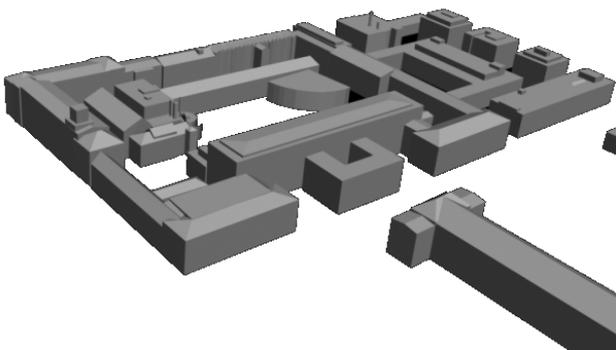


Figure 4: 3d building model with polygon surfaces

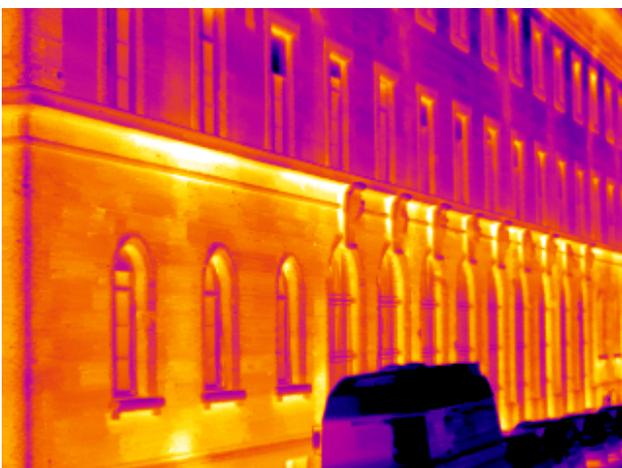


Figure 5: IR image, intensity values are coded as 256 color table

The homography surface estimation leads to a relative camera path and estimated surface for the image sequence (Fig. 6). Using the given GPS coordinates, we can assign each of the estimated surfaces to a polygon surface of the model.

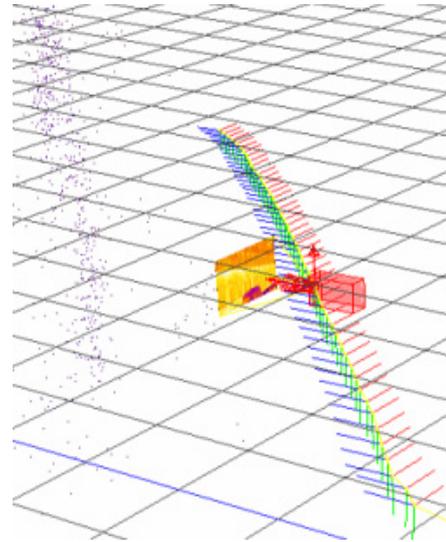


Figure 6: Camera path and point cloud of corresponding points of the image sequence

As mentioned, raycasting is used to determine the visible surface for every pixel and then calculate its texture coordinates. From these texels of one image a partial surface texture for every visible surface is generated. Figure 7 demonstrates the sampling rate and the resolution distribution. To the right, the resolution is decreasing due to the perspective view. Areas of the surface, which could not be seen in one image, are left in black. The black stripes are image pixels, where no texel is transformed to.

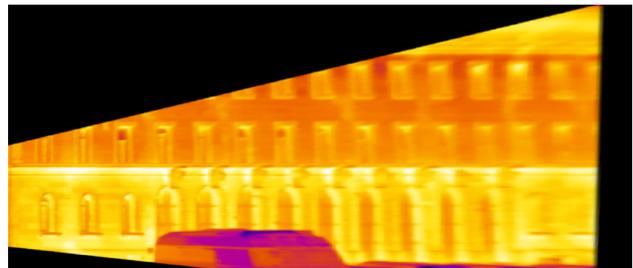


Figure 7: Partial surface texture for the image of Fig. 5, intensity values are coded as 256 color table.

For composing a complete texture for a building façade, for every pixel of the final texture its value is interpolated from the stored partial surface textures. Figure 8 shows a combined surface texture built of twenty partial textures from one sequence in forward view. The black areas from the partial textures are filled from other partial textures. In difference to the partial texture, where the resolution is decreasing from left to right, here the resolution is decreasing from the first floor to the roof and the first floor to the ground, caused by the viewing direction of the camera and the texture combination (see section 3.4 and Fig. 2 and 3).

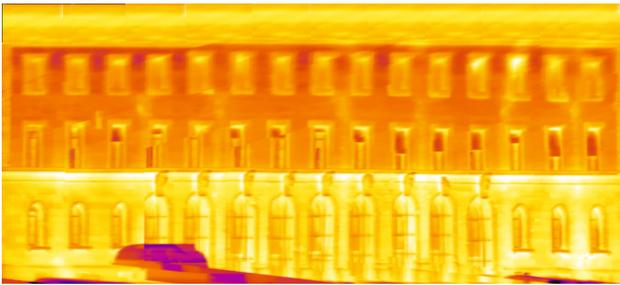


Figure 8: Surface texture generated from twenty partial textures.

For the texture in Fig. 8, no invisible parts remain due to the fact that there is no occlusion for this façade. Other façades are not completely visible from the sequence so that there remain invisible parts left in black (Fig. 9).

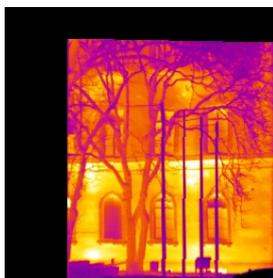


Figure 9: Texture of a partially occluded façade where the black part is not visible in any partial texture for this façade.

5 DISCUSSION

Infrared light has special characteristics that lead to many problems. The physical behavior of the IR spectrum causes camera systems with lower resolution than normal video or photo cameras. The appearance of a façade in the infrared depends on outer conditions like weather, daytime, sun, temperature and inner conditions like building structure or heating. Many things to be seen in one infrared image, can not be seen in another because the conditions have changed and many façade structures that can be seen in visible can not be seen in infrared and vice versa. In addition, the small viewing angle and resolution of the camera does not allow the record of a complete building in dense urban areas with narrow streets in one image. The oblique view caused by the viewing angle and resolution of the cameras leads to self occlusion of buildings. This is partially solved by recording two different views per façade, one in oblique forward view and one in oblique backward view (Fig. 10). But, there are still remaining parts of buildings that are not visible. The given camera parameters for the projection have to be accurate, because the texture combination can only adjust small offsets between the partial textures. Errors in the rotation of the camera lead to completely wrong intersection points and thus wrong surface assignments and texture coordinates. A strategy based on the 5-point algorithm of Nistèr (Nistèr, 2004) will be investigated to deal with discontinuous façades.

In the strategy mentioned in this paper the resolution of the combined final texture is restricted to the resolution of the partial textures from the image sequence because one partial texture is simply overwriting another. In addition, this strategy is only feasible for the combination of textures of one sequence

and with constant view direction. To reach a more general approach the texture coordinates generated during the ray casting from every single image should be combined as texture coordinate 2d point cloud. With this hole set of texture coordinates the calculated combined texture is interpolated from a much more comprehensive basis and thus could reach a higher resolution than the input textures. A necessary precondition for this strategy is the improvement of the determination of the camera parameters to avoid blurred textures from incorrect texture coordinates.

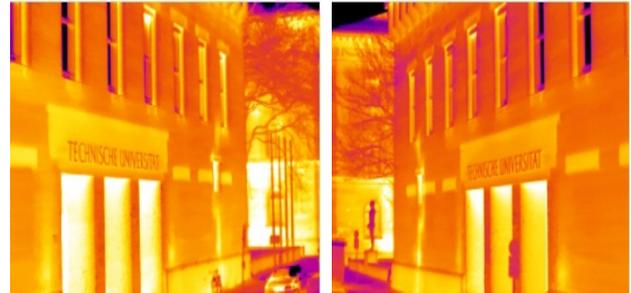


Figure 10: Forward and backward view from two sequences of the same building parts

As seen in figure 8, the windows of the façade can easily be seen and, because the whole façade is in one image, the rows and columns can be identified to search for further windows, which could be detected directly. In further processing steps, for structures like the windows, the intersection points corresponding to the pixels of the window corners can be used to create triangles in the 3d model to represent the window in geometry instead of texture.

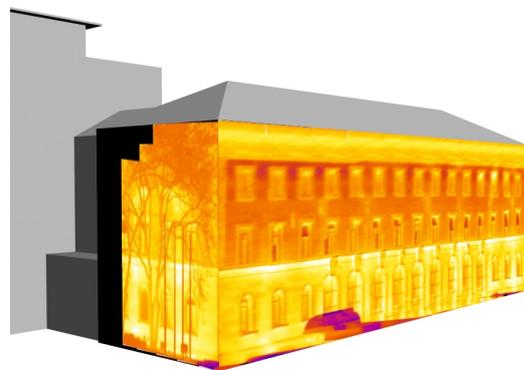


Figure 11: Textures of Fig. 8 and 9 mapped onto the 3d building model of Fig. 4

ACKNOWLEDGEMENTS

The authors thank Dr. Clement, Dr. Schwarz and Mr. Kremer of FGAN-FOM, Department Optronik Systems Analyses, Ettlingen, for equipment and their assistance during the recording campaign.

Our work is part of the DFG (German Research Society) research project “Enrichment and Multi-purpose Visualization of Building Models with Emphasis on Thermal Infrared Data“ (STI 545/1-2) as part of the bundle project “Interoperation of 3D Urban Geoinformation (3DUGI)”.

REFERENCES

- Fischer, M.A., Bolles, R.C., 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, *Communications of the ACM*, vol.24(6), June 1981, pp. 381-395
- Foley, J.D., 1995. *Computer Graphics: Principles and Practice*, Addison-Wesley, ISBN 02-018-4840-6
- Haralick, R.M., Lee, C.N., Ottenberg, K., Nolle, M., 1994. Review and analysis of solutions of the 3-point perspective pose estimation problem, *IJCV*, vol.13(3), pp. 331-356
- Hartley, R.L., Zisserman, A., 2000. *Multiple View Geometry in Computer Vision*, Cambridge University Press, ISBN: 0521623049
- Hinz, S., Stilla, U., 2006. Car detection in aerial thermal images by local and global evidence accumulation, *Pattern Recognition Letter*, vol. 27, pp. 308-315
- Klingert, M., 2006. The usage of image processing methods for interpretation of thermography data “,17th International Conference on the Applications of Computer Science and Mathematics in Architecture and Civil Engineering, Weimar, Germany, 12-14 July 2006
- Koskeleinen, L., 1992. Predictive maintenance of district heating networks by infrared measurement, *Proc. SPIE*, vol. 1682, pp. 89–96
- Lo, C.P., Quattrochi, D.A., 2003. Land-Use and Land-Cover Change, Urban Heat Island Phenomenon, and Health Implications: A Remote Sensing Approach, *Photogrammetric Engineering & Remote Sensing*, vol. 69(9), pp. 1053–1063
- Nichol, J., Wong, M.S., 2005. Modeling urban environmental quality in a tropical city, *Landscape and urban Planning*, vol.73, pp.49-58
- Nistèr, D., 2004. An efficient solution to the five-point relative pose problem, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 6, pp. 756-777, Jun. 2004
- Quan, L., Lan, L.D., 1999. Linear n-point camera pose determination, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.21(8), pp. 774-780
- Quattrochi, D.A., Luvall, J.C., 1999. Thermal infrared remote sensing for analysis of landscape ecological processes: methods and application, *Landscape Ecology*, vol. 14, pp. 577–598
- Siegert, F., Zhukov, B., Oertel, D., Limin, S., Page, S.E., Rielay, O., 2004. Peat fires detected by the BIRD satellite, *International Journal of Remote Sensing*, vol. 25(16), pp. 3221-3230
- Stilla, U., Michaelsen, E., 2002. Estimating vehicle activity using thermal image sequences and maps, Symposium on geospatial theory, processing and applications. *International Archives of Photogrammetry and Remote Sensing*, vol.34,Part 4
- Triggs, B., 1999. Camera pose and calibration from 4 or 5 known 3d points, *Proc. International Conference on Computer Vision (ICCV'99)*

ROBUST DISPARITY MAPS WITH UNCERTAINTIES FOR 3D SURFACE RECONSTRUCTION OR GROUND MOTION INFERENCE

A. Jalobeanu¹ and D. D. Fitzenz²

¹ LSIIT (CNRS - Univ. Strasbourg 1), Illkirch, France
(e-mail: jalobeanu@lsiit.u-strasbg.fr, web: lsiit-miv.u-strasbg.fr/paseo/jalobeanu.php)

² EOST (CNRS - Univ. Strasbourg 1), Strasbourg, France
(e-mail: fitzenz@eost.u-strasbg.fr, web: eost.u-strasbg.fr/dfitzenz)

Commission III/2

KEY WORDS: Optical sensors, Remote sensing, Stereo, Uncertainties, Error maps, Change modeling, Digital terrain models

ABSTRACT:

Disparity maps estimated using computer vision-derived algorithms usually lack quantitative error estimates. This can be a major issue when the result is used to measure reliable physical parameters, such as topography for instance. Thus, we developed a new method to infer the dense disparity map from two images. We use a probabilistic approach in order to compute uncertainties as well. Within this framework, parameters are described in terms of random variables. We start by defining a generative model for both raw observed images given all model variables, including disparities. The forward model mainly consists of warping the scene using B-Splines and adding a radiometric change map. Then we use Bayesian inference to invert and recover the a posteriori probability density function (pdf) of the disparity map. The main contributions are: The design of an efficient fractal model to take into account radiometric changes between images; A multigrid processing so as to speed up the optimization process; The use of raw data instead of orthorectified imagery; Efficient approximation schemes to integrate out unwanted parameters and compute uncertainties on the result. Three applications could benefit from this disparity inference method: DEM generation from a stereo pair (along or across track), automatic calibration of pushbroom cameras, and ground deformation estimation from two images at different dates.

1 INTRODUCTION

Computational stereo vision aims at matching pixels from a stereo image pair; a great number of methods have already been developed – see (Brown et al., 2003) for a review. A few calculate dense disparity maps. The majority work along scanlines, assuming images have been rectified beforehand using epipolar geometry to restrict the search space to a single dimension (e.g. horizontal displacements). However, we claim that resampled images are not suitable for a proper probabilistic inference. We advocate the processing of raw data instead, because resampling destroys the independence properties of the noise by creating correlations, whereas only original pixels can be assumed independent variables. Therefore we have to estimate disparities in 2D while most advanced techniques only work in 1D; this is the only way of preserving the integrity of the data and deriving non-biased estimators. Experiments have shown that resampling via classical interpolation (whether bilinear or bicubic) induces systematic errors that can not be neglected when aiming at a sub-pixel accuracy. Working with raw data shall then result in increased accuracy.

The new approach provides a quantitative measure of uncertainty while most methods only compute ad-hoc matching or correlation quality measures. Disparity maps with uncertainties have been computed (Blake et al., 2003), however this was made possible by working in 1D. When it comes to robustness to illumination changes, probabilistic models have been proposed (Zhang et al., 2006) that rely on an illumination ratio map; this was also possible by using 1D disparities. Nonparametric approaches based on mutual information (Hirschmuller, 2005) were developed to handle such changes without having to explicitly model them; however it is unclear how they can be extended to account for spatially adaptive changes due to the terrain reflectance without resorting to a tremendous number of parameters to store the required histograms. In our case, robustness to illumination is achieved via probabilistic change modeling in 2D and does not require extra parameters. Techniques based on correlations, robust or not, usu-

ally rely on the choice of a window size which is often arbitrary, and those involving smoothness priors are often fine tuned manually; our technique is designed to be entirely automated as neither window size nor smoothness parameters need to be adjusted.

2 THE FORWARD MODEL

2.1 Deterministic part: rendering

We define the sets of bidimensional indices related to displacement parameters and pixels respectively as $\Omega_\theta = \{1 \dots n_x\} \times \{1 \dots n_y\}$ and $\Omega_p = \{1 \dots N_x\} \times \{1 \dots N_y\}$. There are $N = N_x N_y$ pixels and $n = n_x n_y$ disparity parameters.

The disparity map d is a set of 2D vectors that define a mapping from image 1 to image 2; we use a B-Spline representation (Thévenaz et al., 2000) at scale R parametrized by the coefficients Δ . The disparity can be set at a coarser scale than the image, so as to allow for a multigrid optimization scheme as shown in section 4.2. A coverage factor $\alpha \in [0, 1]$ is defined for each pixel, describing the total contribution of the motion model to the local displacement d , and enabling us to weight each data pixel according to the ability of the model to predict that pixel.

$$d_p = \sum_{j \in \Omega_\theta} \frac{1}{\alpha_p} w_{pj} \Delta_j \quad \text{for } \alpha_p > 0 \quad (1)$$

$$\text{where } w_{pj} = \varphi\left(\frac{1}{R}p - j\right) \quad \text{and } \alpha_p = \sum_{p \in \Omega_p} w_{pj} \quad (2)$$

The 2D kernel φ is separable, i.e. $\varphi(v) = s(v^x) s(v^y)$ where s is the 1D B-Spline 3 kernel. We define the sampled disparities $D_j = d_{Rj}$, so that we have $D = S\Delta$ where S is the convolution operator achieving the interpolation for discrete spatial positions (discrete filter $\varphi(p)$, $p \in \Omega_p$).

The image X^1 is the reference for displacement, i.e. $X^1 \equiv X$, whereas the image X^2 is obtained by warping X through the

disparity map using B-Spline interpolation, which is denoted by $W^\Delta(X)$. This approximation remains valid as long as the mapping from image 1 to image 2 preserves the area. This is generally the case for disparity maps since the displacements can be seen as local shifts, and the global scaling is rather negligible, especially in along-track stereo. If we denote the Spline coefficients of X by L^0 such that $L^0 = S^{-1}X$, we have:

$$X_p^2 = W^\Delta(X)_p = \sum_{k \in \Omega_p} L_k^0 \varphi(p - k - d_p) \quad (3)$$

2.2 Self-similar change modeling

The observed images are denoted by Y^1 and Y^2 . Image X^1 is assumed to be the reference, whereas X^2 undergoes a radiometric transformation which models reflectance effects, shadows, and relatively large-scale changes that can occur between multi-date observations. We assume all these structured radiometric changes can be embedded in an additive term C . In this work, we also assume that C contains the observation noise. To further simplify the modeling, the noise from both images is embedded in C while the reference image remains noise-free:

$$Y^1 = X^1 \quad \text{and} \quad Y^2 = X^2 + C \quad (4)$$

One could argue that due to surface physics, multiplicative transforms are more likely; in fact, to be really accurate, one should take into account at least 3 phenomena – a multiplicative field related to non-Lambertian reflectance properties, clouds and atmospheric transparency, an additive field related to ambient lighting and atmospheric scattering, and also a nonlinear transfer function that is not necessarily shift invariant. Multidate changes are even more complex to describe. Therefore a simple additive term C shall be a reasonable choice in practice, and we will also assume X^2 and C are independent in the method proposed here.

Change maps can be modeled in various ways. Among all possible models, few are simple enough to allow for an efficient implementation. We want to take into account two essential properties: power spectrum decay (high frequency changes are less likely than the low frequency ones) and spatial adaptivity (there is no reason all areas behave similarly, as there might be occlusions, shadows or more dramatic changes). We propose to use a Gaussian fractal process, motivated by the self-similarity of natural images (Jalobeanu et al., 2003). In the frequency space (where spatial frequencies are denoted by u and v), all coefficients are independent Gaussian variables whose variance obey a self-similar law defined by an energy ε^2 and an exponent q :

$$P(\mathcal{F}[C]_{uv}) = \mathcal{N}(0, \varepsilon^2 (u^2 + v^2)^{-q}) \quad \text{for } u, v \neq 0, 0 \quad (5)$$

We will assume a fractal exponent equal to 1 which is a commonly encountered value in natural scenes. In order to remain in the image space we use Markov Random Fields; this is required by the spatial adaptivity, best parametrized in the image space and not in the frequency space. The expression above corresponds to a diagonal inverse covariance equal to $\varepsilon^{-2} (u^2 + v^2)$ in the Fourier space. Its counterpart in the image space is proportional to the linear operator H :

$$H = G_x^t G_x + G_y^t G_y \quad (6)$$

where G_x and G_y denote the image intensity gradients (finite differences) in the x and y directions. This is all made possible by setting $q = 1$ (non-integer values require fractional derivatives, whose computational complexity is considerably higher). We set the gradients to zero on the boundaries. We get:

$$P(C | \lambda) = \frac{1}{Z_\lambda} e^{-\lambda (AC)^t H AC} \quad (7)$$

where λ is a smoothness parameter (related to the factor ε) and Z_λ the corresponding normalization coefficient. Using bounded gradient values enables us to define a proper distribution.

In (7), A is a diagonal matrix that helps achieve spatial adaptivity. The variance of C is divided by A . The simplest choice consists of setting $A_p^2 = \alpha_p$ so as to take into account the coverage factor α_p for each pixel p . Indeed, this factor allows us to put a weight on each data pixel according to the ability of the model to predict this pixel value. Any pixel not related to the model (i.e. that the model can not explain) shall have a weight equal to zero so as to cancel its contribution to the model, which is achieved by an infinite variance, thus allowing the change map value C_p to grow arbitrarily large. Conversely, pixels fully covered by the model ($\alpha=1$) are not affected.

Other pixel-dependent weights could be included here, such as a mask allowing to exclude some of the pixels (classified as unreliable beforehand) from the inference. The weights could also be updated recursively by analyzing the change map after each inference step to make the method more robust.

2.3 A prior disparity map model

To complete the forward model we have yet to define a prior model for the unknown disparity map. A simple choice is a smoothness prior having the same fractal properties as the change model, particularly suitable for disparities related to 3D surfaces or natural phenomena. To keep the approach as general as possible we do not consider the epipolar constraint as in (Deriche et al., 1994) and we set two separate smoothness parameters ω_x , ω_y for disparities along each direction.

$$P(\Delta | \omega) = \frac{1}{Z_{\omega_x} Z_{\omega_y}} e^{-\omega_x (\Delta^x)^t H^*(\Delta^x) - \omega_y (\Delta^y)^t H^*(\Delta^y)} \quad (8)$$

Here Z_{ω_x} , Z_{ω_y} denote normalization coefficients. The operator H^* can be different from H . In principle, the model should apply to the disparities D , not to its Spline coefficients Δ . Therefore we decided to choose $H^* = S^t H S$ in order to use a self-similar prior similar to the change map prior.

We will discuss later how to estimate the smoothness parameters. There is no immediate need for a prior pdf since these parameters are substantially overdetermined. They are related to the underlying structure of the motion field, a special case being along-track stereo where the displacement map along x is particularly smooth, in which case very large values of ω_x are expected.

2.4 Forward model and the related Bayesian network

We just defined the elements of a full generative model describing the formation of an image pair from a single scene X , given a deformation field Δ , a change map C and a noise variance map σ^2 as well as their respective parameters. A Bayesian network displays all causality relations (see Fig. 1) and enables us to write the joint pdf of all model variables as a product of prior pdfs and conditional pdfs. The conditional pdf of the variables given the data is proportional to this joint pdf, since the data is fixed.

3 THE INVERSE PROBLEM

Now we need to solve the inverse problem (with respect to the direct model just defined). In the Bayesian approach (Gelman et al., 1995), we aim at the determination of the a posteriori pdf, i.e. the pdf of the variables of interest Δ given the data Y^1 , Y^2 , written as $P(\Delta | Y^1, Y^2)$. Three major steps are involved:

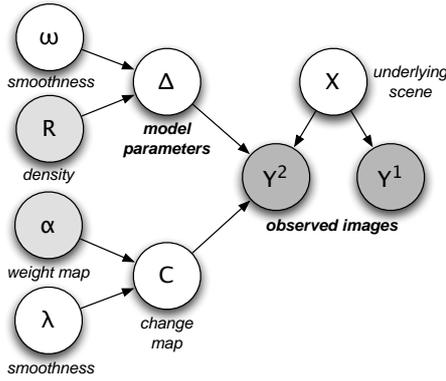


Figure 1: Bayesian network related to the proposed forward model, describing the formation of the observed images given all model variables. Nodes with converging arrows represent conditional pdfs, while all other nodes represent priors. Shaded nodes are fixed (or observed) variables.

- Marginalization, to formally integrate the joint pdf with respect to all variables that are not of interest (also known as nuisance variables), denoted by $\Theta = \{X, C, \lambda, \omega_x, \omega_y\}$ such that: $P(\Delta | Y^1, Y^2) \propto \int P(Y^1, Y^2, \Delta, \Theta) d\Theta$
- Optimization, which aims at finding the values of Δ that maximize the posterior pdf (this step might involve the optimization of other quantities as required by marginalization). In practice an energy function U , defined as the -log of the marginal posterior pdf, is being minimized.
- Gaussian approximation around the optimum to determine parametric uncertainties on the result.

3.1 Image resampling and change map marginalization

The observed image Y^2 is now fixed, so we set $z^2 \equiv Y^2$ and $z^1 = W^\Delta(Y^1)$ since we assumed $Y^1 \equiv X$. If we denote by L the B-Spline coefficients of Y^1 , we have:

$$z_p^1 = W^\Delta(Y^1)_p = \sum_{k \in \Omega_p} L_k \varphi(p - k - d_p) \quad (9)$$

The difference $z^{12} = z^2 - z^1$ defined this way is equal to the radiometric changes C , according to Eqn. (4). See Fig. 3 for an illustration. If we had a white Gaussian observation noise instead of the structured change model (by replacing H with identity), the obvious procedure would consist of minimizing the classical sum of squared differences (SSD) between Y^2 and $W^\Delta(Y^1)$. In general, this SSD-based procedure fails as the changes are spatially structured (hence the use of normalized SSD by some authors, but without explicit radiometric changes). Since we have two deterministic relations (4), integrating out X and C gives

$$P(Y^1, Y^2 | \Delta, \lambda) \propto \frac{1}{Z_\lambda} e^{-\lambda (Az^{12})^t H Az^{12}} \quad (10)$$

3.2 Estimating the change smoothness parameter

A noninformative, parameter-free prior is assumed for λ since we have little knowledge about the changes; discussing how to choose the prior pdf is beyond the scope of this paper. The integration with respect to λ can be done via the Laplace approximation as explained in (MacKay, 2003), which amounts to evaluating a Gaussian integral. The approximation consists of considering that the -log of the integrand is a quadratic form, defined by an minimum (location and value) and a curvature at the minimum. As there is a single parameter for all pixels, the curvature is very high, therefore the integrand behaves like a Dirac function so this approximation is valid. To evaluate the integral this way, we need

to calculate the value at the optimum $\hat{\lambda}$ (the location itself does not matter) as well as the second derivative $-\partial^2 \log P / \partial \lambda^2$; the latter does not depend on Δ so it simply acts as a constant factor and is ignored. The optimum is given by

$$\hat{\lambda} = \frac{N}{2 \Phi(Az^{12})} \quad \text{where} \quad \Phi(X) = X^t H X \quad (11)$$

We finally get the integrated likelihood for Δ :

$$P(Y^1, Y^2 | \Delta) \propto e^{-\frac{N}{2} \log \Phi(Az^{12})} \quad \text{with} \quad \Phi(Az^{12}) \neq 0 \quad (12)$$

We never have $\Phi = 0$ if $R > 1$, because z^1 and z^2 can not be equal for all pixels since the images are noisy and $R > 1$ prevents any over-fitting (there is more than one pixel per parameter).

3.3 Estimating the disparity smoothness

Instead of using the smoothness prior (8) with fixed values of the parameters ω , which would require a rather complex estimation procedure to find optimal values (Jalobeanu et al., 2002), we can use an integrated prior by integrating out ω_x and ω_y from the very beginning. This way we lose the benefits of having a quadratic regularization energy but we make the prior parameter-free. It also makes the full inference method parameter-free since λ was already integrated out.

The pdf (8) can be approximated by a Gaussian around its mode as we did in the previous paragraph, so we apply exactly the same reasoning as we did with λ (noninformative prior, and Laplace approximation), which yields:

$$P(\Delta) \propto e^{-\frac{n}{2} \log \Phi^*(\Delta^x) - \frac{n}{2} \log \Phi^*(\Delta^y)} \quad (13)$$

This enables us to finally express the posterior $P(\Delta | Y^1, Y^2)$, proportional to the product of expressions (12) and (13):

$$P(\Delta | Y^1, Y^2) \propto e^{-\frac{N}{2} \log \Phi(Az^{12}) - \frac{n}{2} \log \Phi^*(\Delta^x) - \frac{n}{2} \log \Phi^*(\Delta^y)} \quad (14)$$

A degenerate solution could be found for $\Phi^* = 0$, since nothing prevents us from having a perfectly smooth disparity map (for instance Δ^x is very smooth in along-track stereo, almost linear in our tests as shown in the result section on Fig. 4). The interpretation of the posterior pdf (14) needs special care: we should not seek the global optimum, but rather the most 'significant' one. The peak related to degenerate solutions is the highest (in fact, this posterior is improper, with a singularity at $\Phi^* = 0$, but it can be seen as the limit of a sequence of proper pdfs). However it can be shown by computing the second derivatives that this peak is very narrow compared to the one related to the acceptable solution, so that in terms of probability integrated over the corresponding volume, a meaningful, non-degenerate solution is more probable – even if does not maximize the pdf.

4 PROPOSED ALGORITHM

4.1 A fully unsupervised method

In order to maximize the posterior (14), the energy U needs to be minimized with respect to the disparity map parameters Δ :

$$U = \frac{1}{2} (N \log \Phi(Az^{12}) + n \log \Phi^*(\Delta^x) + n \log \Phi^*(\Delta^y)) \quad (15)$$

As mentioned above, special care has to be taken to avoid the potential well around the singularity. This can be effectively done by replacing $\log(\Phi^*)$ by $\log(\delta^2 + \Phi^*)$ where δ is strictly positive to avoid the singularity. When the solution is found, δ can be

set to zero, hoping we escaped the sphere of influence of the singularity and will now converge to the acceptable solution. One can also adopt a more careful scheme inspired from graduated non convexity where δ starts large and is progressively decreased until it reaches zero, thus ensuring convergence.

The marginalization that helps make the problem parameter-free does not add much difficulty to the optimization which is already nonlinear. Indeed, z^{12} has a nonlinear dependence on Δ because of the resampling process. The energy (15) is not quadratic (and also not convex). Thus, solving this problem needs special handling to avoid local minima and find the desired solution. Simple Newton-like methods can not be used, and we must resort to a nonlinear gradient descent. This can be sensitive to the initialization, hence the multigrid technique discussed in section 4.2.

We choose to use a nonlinear conjugate gradient method (Press et al., 1993), which only requires the first derivatives of U . The derivatives of the data term are denoted by ζ^x :

$$\zeta_j^x = \frac{\partial}{\partial \Delta_j^x} \frac{N}{2} \log \Phi(Az^{12}) = \frac{N}{\Phi(Az^{12})} (HAz^{12})^t g_j^x \quad (16)$$

where g_j^x is related to the derivative of z^{12} and is defined as:

$$(g_j^x)_p = -\alpha_p \frac{\partial z_p^{12}}{\partial \Delta_j^x} = w_{pj} W_{\partial_x} \Delta_j^x (Y^1)_p \quad (17)$$

Here $W_{\partial_x}^{\Delta}$ denotes an interpolation based on the derivative of φ , involving the kernel $\varphi'^x(v) = s'(v^x) s(v^y)$ instead of φ :

$$W_{\partial_x}^{\Delta} (Y^1)_p = \sum_k L_k \varphi'^x(p - k - d_p) \quad (18)$$

which is computed at the same time as the usual interpolation W^{Δ} . We get similar equations for Δ^y . The derivatives of the full energy (data term and prior) are finally given by:

$$\frac{\partial U}{\partial \Delta_j^x} = \zeta_j^x + \frac{n}{\Phi^*(\Delta^x)} H^* \Delta_j^x \quad (19)$$

4.2 Multigrid optimization

The most classical approach to multigrid is to start with a coarse disparity model (large R), perform the optimization, and refine recursively by initializing finer scale models with the previous, coarser scale. This has been applied to image registration in (Thévenaz et al., 1998). A dyadic scheme is generally used, starting with $R = R_0$ and dividing by 2 at each step. Coarse estimates are discarded, since they are only used to initialize the next scale.

However, when processing large size images, one had better estimate coarse deformation models from subsampled versions of the input images rather than from the full size images, especially at very coarse scales. A Spline pyramid (Unser et al., 1993) is computed for both images Y^1 and Y^2 . At each scale, we apply the inference procedure described above, with a fixed value of R ; the range $2 \leq R \leq 8$ achieves a good trade-off between model density and number of data points per parameter. Then, the model is refined through Spline subdivision since the parameters Δ are actually the Spline coefficients of the dense disparity field d . The refined model is used to initialize the next scale.

4.3 Computing uncertainties

The inverse covariance matrix related to a Gaussian approximation of the posterior pdf around the optimum Δ is defined by 4 blocks, related to second derivatives with respect to Δ^x and Δ^y :

$$\Sigma_{\Delta}^{-1} = \left(\begin{array}{cc} \frac{\partial^2 U}{\partial \Delta^x \partial \Delta^x} & \frac{\partial^2 U}{\partial \Delta^x \partial \Delta^y} \\ \frac{\partial^2 U}{\partial \Delta^y \partial \Delta^x} & \frac{\partial^2 U}{\partial \Delta^y \partial \Delta^y} \end{array} \right)_{\Delta=\hat{\Delta}} \quad (20)$$

Using the first derivatives previously calculated (16)-(18) and assuming a locally linear resampling process W^{Δ} , we get:

$$\frac{\partial^2 U}{\partial \Delta_k^x \partial \Delta_l^x} \simeq \frac{N}{\Phi(Az^{12})} (g_k^x)^t H(g_l^x) - \frac{1}{N} \zeta_k^x \zeta_l^x + \frac{n}{\Phi^*(\Delta^x)} H^* \quad (21)$$

where all expressions are evaluated at $\Delta = \hat{\Delta}$. We obtain a similar expression for Δ_k^y and Δ_l^y . The cross-terms are:

$$\frac{\partial^2 U}{\partial \Delta_k^x \partial \Delta_l^y} \simeq \frac{N}{\Phi(Az^{12})} (g_k^x)^t H(g_l^y) - \frac{1}{N} \zeta_k^x \zeta_l^y \quad (22)$$

The quadratic form is $u^t H v = (G_x u)^t (G_x v) + (G_y u)^t (G_y v)$.

For most applications, we need uncertainties on the disparities $D = S \Delta$ (in practice the optimization is simpler to perform with respect to Δ , this is why we do not use D in the first place). If we need Σ_D^{-1} rather than Σ_{Δ}^{-1} , the following expression can be used for the conversion:

$$\Sigma_D^{-1} = (S^{-1})^t \Sigma_{\Delta}^{-1} (S^{-1}) \quad (23)$$

where S^{-1} is the operator that transforms a vector into a series of Spline coefficients, which can be implemented very efficiently (Thévenaz et al., 2000). Notice that this is only needed once the optimization procedure has been completed, therefore does not affect the computational cost of the disparity estimation itself.

In principle, uncertainties are expressed through variances and covariances, which have a physical meaning: the former directly relates to confidence intervals on the estimated parameters, and the latter give the correlation between model variables. However, the matrix (20) needs to be inverted, which is difficult in practice because of its size ($2n \times 2n$). Therefore we approximate covariances between neighboring disparity parameters by neglecting long-range interactions.

For each covariance to be computed between variables indexed by i and j , we select a small block of the matrix Σ^{-1} by picking only the entries for variables directly connected to i and j . Obviously, in the inversion, variables that do not interact shall not be involved in the computation. In practice, for a given spatial position j (diagonal elements of each block of Σ , related to Δ_j^x and Δ_j^y) if we restrict to the 8 nearest neighbors $j + (\pm 1, \pm 1)$ for both Δ^x and Δ^y , we only need to invert a 18×18 matrix. With 4 nearest neighbors the size of the matrix reduces to 10×10 . This has to be repeated for each spatial location k . Notice that efficient iterative optimization techniques can advantageously replace matrix inversion (Jalobeanu and Gutiérrez, 2007).

If uncertainties do not need to be interpreted, but rather stored and propagated through to other processing algorithms, the inversion can be avoided: only the inverse covariance matrix needs to be propagated. If there are too many terms it can be simplified so as to limit the redundancy of the end result; for instance one can provide inverse covariances related to the 4 nearest neighbor Δ variables, which only requires to store and propagate 7 extra terms for each location j (in addition to the estimates $\hat{\Delta}_j^x$ and $\hat{\Delta}_j^y$); refer to table 1 for details.

Self	$\Sigma_{j,j}^{-1} \text{ } ^{xx}, \Sigma_{j,j}^{-1} \text{ } ^{yy}$
Cross (xy)	$\Sigma_{j,j}^{-1} \text{ } ^{xy}$
Horizontal	$\Sigma_{j+(1,0),j}^{-1} \text{ } ^{xx}, \Sigma_{j+(1,0),j}^{-1} \text{ } ^{yy}$
Vertical	$\Sigma_{j,j+(0,1)}^{-1} \text{ } ^{xx}, \Sigma_{j,j+(0,1)}^{-1} \text{ } ^{yy}$

Table 1: Uncertainty terms (inverse covariances, limited to 4 nearest neighbor interactions) produced by the proposed disparity inference algorithm, after inverse covariance simplification.

5 PRELIMINARY RESULTS AND DISCUSSION

5.1 Tests on real Mars Express images

We chose a test area extracted from a raw panchromatic, along-track stereo pair taken by the Mars Express HRSC instrument (ESA). The image was downsampled by a factor 8 using a Spline pyramid (preserving noise statistics) to ensure small displacements (<10 pixels) then a study area was selected ($N = 64 \times 64$, see Fig. 2). The initial disparity maps were $D_x =$ linear function of x and $D_y = 0$. A radius $R = 4$ was chosen for the model resolution ($n = 16 \times 16$). Convergence was reached in less than 50 iterations for this particular initialization; a multigrid strategy would significantly reduce this number, however the purpose of the test is to check the validity of the proposed energy functional (15) rather than study how to minimize it most efficiently.

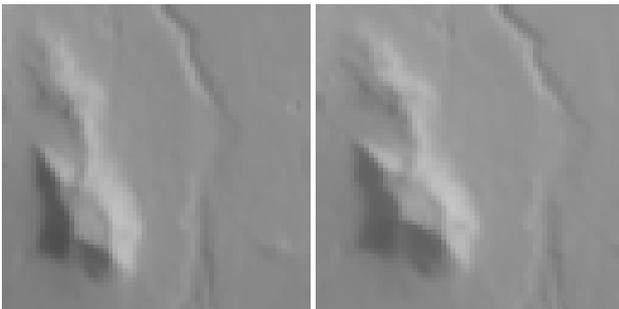


Figure 2: Left: Y^1 ; right: Y^2 . ESA Mars Express HRSC Orbit 0905 (stereo s12/s22), subsampling factor 8, extracted region 64×64 pixels.

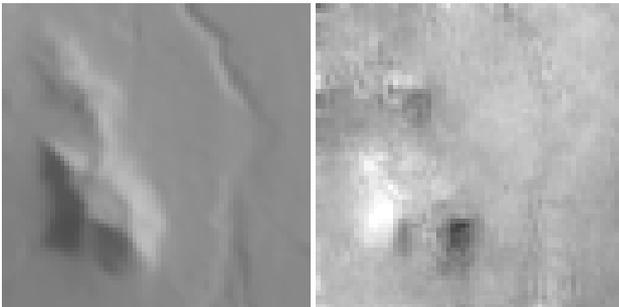


Figure 3: Left: $W^\Delta(Y^1)$ (Y^1 warped using the estimated disparity map); right: corresponding change map $C = Y^2 - W^\Delta(Y^1)$, contrast enhanced (factor 10) to display the effects of non-Lambertian reflectance.

Fig. 4 displays the estimated disparities D_x and D_y ; obviously D_x is unrelated to the topography and appears very smooth. Uncertainties also shown (bottom Fig. 4) are preliminary results; we only show the inverse of the diagonal of Σ_D^{-1} without the cross-terms, however these preliminary error maps already carry valuable information that can be interpreted as error bars on D_x and D_y (standard deviation). It clearly illustrates the spatial variability of errors; the higher the contrast of radiometric features (e.g. texture, edges), the lower the uncertainty. It falls below 0.1 pixel as long as there are enough details, and well below 0.05 pixel near edges, whereas it can reach 0.3 pixel in the smoothest areas. Notice that these estimates do not depend on the observation noise parameters, which are unknown, nonetheless they exploit the available statistics throughout the inference procedure. Notice also that there is no available ground truth for disparities.

5.2 Application to 3D reconstruction

If the imaging geometry is known and reliable (e.g. attitude and position of the satellite computed from the metadata), then 2D displacements can be directly converted into 1D heights, which amounts to projecting the 2D pdfs of disparities to get 1D pdfs

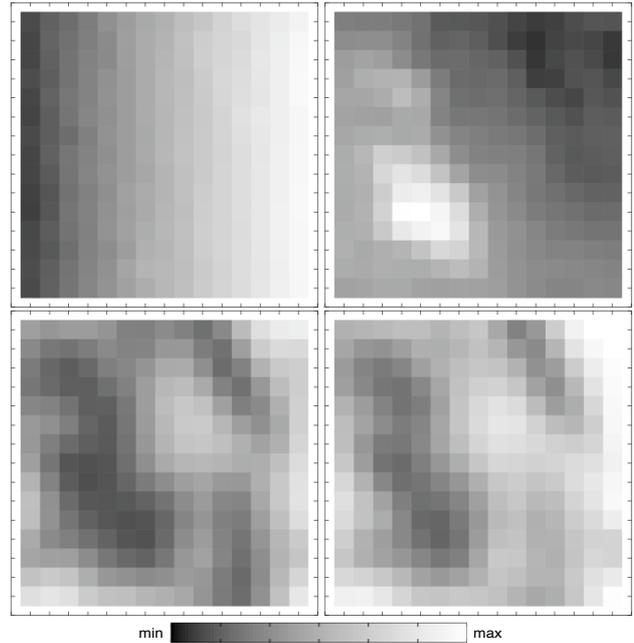


Figure 4: Top: estimated disparities (left: D_x , range=[0,9], right: D_y , range=[-3.5,-1.5]); bottom: related standard deviations, range=[0,0.3].

of elevations. This will provide a DEM as a height field (longitude, latitude and elevation) where the sampling corresponds to the pixels of image Y^2 , most probably irregular on the ground. Therefore it needs to be resampled on a regular grid (again we recommend B-Spline interpolation), and the uncertainties need to be converted accordingly using an equation similar to (23). The same process applies to pixel values from both input images in order to achieve orthorectification, which amounts to providing a textured terrain model using reflected radiance maps. As opposed to traditional orthorectification techniques, this should be done using probability theory, which produces uncertain and correlated pixel values as the estimated elevations are also uncertain. Moreover, the resampling may also produce a blur due to the geometric uncertainty related to probabilistic elevations.

5.3 Application to automatic camera calibration

Remote sensing images are acquired with pushbroom systems whose parameters are not always well-known. Even if a reasonably accurate calibration can be achieved through star-based optical navigation or using onboard sensors, applications aiming at a sub-pixel precision can not rely on it. Indeed, the orbital motion is affected by high-frequency vibrations whose parameters are unknown in general. Imperfect trajectory and attitude models derived from the metadata lead to systematic errors in 3D reconstruction that are often inconsistent with the recorded data.

Why not use the data directly without making complex calculations involving a geometry that is not well constrained? If a dense disparity map D is provided as well as the related errors and correlations summarized by Σ_D^{-1} , it should be possible to recover relative camera motion without any other source of information. This has been achieved for approximate, linear camera models (Gupta and Hartley, 1997). Once the relative motion model has been inferred (parameter values as well as their uncertainties can be computed using a probabilistic approach), a relative DEM can be reconstructed in the camera space. It then needs to be converted into the world space using ground control points; however, even without such absolute knowledge, the relative DEM is still a valuable product that makes use of all the radiometric information contained in the stereo pair and its quality is not dependent on possibly misestimated calibration parameters.

While relative camera calibration for rigid detector arrays has been thoroughly investigated in the field of stereo vision (Deriche et al., 1994), the application to pushbroom sensors is still an open research area, especially from the probabilistic point of view.

5.4 Application to ground deformation

There are two main types of ground deformation that would require precise monitoring: gravitational motions (e.g., landslides) and tectonic-driven motions (e.g., co- and/or post-seismic deformation fields). As reviewed in (Delacourt et al., 2004), currently, most of the techniques for monitoring landslide displacement are derived from measurements of reference stations (triangulation, tachometry and GPS measurements). The database of movement provided by these techniques is available only for major landslides for a time span not exceeding 20 years for laser measurements and less than 15 years for GPS. Moreover, due to spatial and temporal heterogeneities of the displacements, such ground based measurements are not sufficient to describe fully the velocity field of a landslide. Remote sensing imagery is a powerful tool for landslide monitoring because it offers a synoptic view that can be repeated at different time intervals. This is also the case for very localized slow tectonic motion. Differential SAR interferometry (DINSAR) has shown its capability for deriving high accuracy maps (at centimeter level) of landslide displacement (Fruneau et al., 1996). However, this technique is affected by severe geometrical and environmental limitations (e.g., loss of coherence due to vegetation). Moreover, the SAR image database is limited to 1991, and later.

In recent years, new techniques based on the correlation of satellite optical images for the processing of deformation maps have been developed. Those techniques have been successfully applied to the measurement of coseismic deformation and complement InSar techniques in particular close to the fault surface trace where no interferogram can be computed. The same techniques were applied to the long-term monitoring of landslides by combining aerial images acquired at different times and very high resolution satellite images (QuickBird). Using a pair of SPOT panchromatic images, (Van Puymbroeck et al., 2000) showed that sub-pixel correlation could provide fault slip measurements with an accuracy of 0.1 pixel (1 meter) (Michel and Avouac, 2002). The approach of (Van Puymbroeck et al., 2000) first consists in resampling the SPOT images into a map projection so that the remaining image pixel offsets are only due to the earthquake ground deformation. In a second step, the coseismic offset map and its error estimate are computed from the phase shift of the Fourier transform of a sliding window. In (Binet and Bollinger, 2005), the method was adapted to SPOT 5 images and the correlation window size used was 256 pixels and the window step was 64 pixels in both directions. Attempts in reducing the window size lead to a noisy offset map near the fault because of the temporal decorrelation and of the low contrast of the ground.

What we want with our approach is 1) to break free from the initial assumption of a rigid motion within each window, 2) to take into account the changes, 3) to compute covariance maps. Indeed, this is required to combine the analysis of real surface deformation information and probabilistic geophysical modeling.

6 CONCLUSIONS AND FUTURE WORK

Our long-term goal can be described as fully automated probabilistic Digital Terrain Model (DTM) generation from uncalibrated stereo pairs (possibly from more than two images). This is part of the SpaceFusion project, funded by the French Research

Agency (ANR); its goal is to combine multiple data to infer the topography, then to fuse the radiometry information into a well-sampled, single reflectance map. A valuable by-product of this project is the determination of 3D deformation fields with error maps if the 3D reconstruction is performed from stereo pairs taken at different dates. This way one can measure the ground motion or the evolution of the topography for analysis or monitoring purposes. To efficiently handle occlusions and abrupt changes, the forward model will need to be extended to allow for spatially adaptive noise statistics, for a better robustness.

REFERENCES

- Binet and Bollinger, 2005. Horizontal coseismic deformation of the 2003 Bam (Iran) earthquake measured from SPOT-5 THR satellite imagery. *Geophys. Res. Lett.* 32(L02307), pp. doi:10.1029/2004GL021897.
- Blake, A., Torr, P., Cox, I. and Criminisi, A., 2003. Estimating uncertainty in dense stereo disparity maps. Microsoft Research Report MSR-TR-2003-93.
- Brown, M., Burschka, D. and Hager, G., 2003. Advances in computational stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28(8), pp. 993–1008.
- Delacourt, C., Allemand, P., Casson, B. and Vadon, H., 2004. Velocity field of the 'La Clapière' landslide measured by the correlation of aerial and QuickBird satellite images. *Geophys. Res. Lett.* 31(L15619), pp. doi:10.1029/2004GL020193.
- Deriche, R., Zhang, Z., Luong, Q.-T. and Faugeras, O., 1994. Robust recovery of the epipolar geometry for an uncalibrated stereo rig. *Lecture Notes in Computer Science* 800, pp. 567–576.
- Fruneau, B., Achache, J. and Delacourt, C., 1996. Observation and modeling of the saint-etienne-de-tinee landslide using SAR interferometry. *Tectonophysics* 265(3-4), pp. 181–190.
- Gelman, A., Carlin, J., Stern, H. and Rubin, D., 1995. *Bayesian Data Analysis*. Chapman & Hall.
- Gupta, R. and Hartley, R. I., 1997. Linear pushbroom cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(9), pp. 963–975.
- Hirschmuller, H., 2005. Accurate and efficient stereo processing by semi-global matching and mutual information. In: *Proc. of CVPR*, pp. II: 807–814.
- Jalobeanu, A. and Gutiérrez, J., 2007. Inverse covariance simplification for efficient uncertainty management. In: *27th MaxEnt workshop, AIP Conference Proceedings*, Saratoga Springs, NY.
- Jalobeanu, A., Blanc-Féraud, L. and Zerubia, J., 2002. Hyperparameter estimation for satellite image restoration using a MCMC Maximum Likelihood method. *Pattern Recognition* 35(2), pp. 341–352.
- Jalobeanu, A., Blanc-Féraud, L. and Zerubia, J., 2003. Natural image modeling using complex wavelets. In: *Wavelets X, SPIE Symposium*, San Diego, CA.
- MacKay, D., 2003. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press.
- Michel, R. and Avouac, J., 2002. Deformation due to the 17 August Izmit, Turkey, earthquake measured from SPOT images. *J. Geophys. Res.*
- Press, W., Teukolsky, S., Vetterling, W. and Flannery, B., 1993. *Numerical Recipes in C: The Art of Scientific Computing*. 2nd edn, Cambridge University Press.
- Thévenaz, P., Blu, T. and Unser, M., 2000. Interpolation revisited. *IEEE Trans. on Med. Imaging* 19(7), pp. 739–758.
- Thévenaz, P., Ruttimann, U. and Unser, M., 1998. A pyramid approach to subpixel registration based on intensity. *IEEE Transactions on Image Processing* 7(1), pp. 27–41.
- Unser, M., Aldroubi, A. and Eden, M., 1993. The L_2 -polynomial spline pyramid. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15(4), pp. 364–379.
- Van Puymbroeck, N., Michel, R., Binet, R., Avouac, J.-P. and Taboury, J., 2000. Measuring earthquakes from optical satellite images. *Applied Optics* 39(20), pp. 3486–3494.
- Zhang, J., McMillan, L. and Yu, J., 2006. Robust tracking and stereo matching under variable illumination. In: *Proc. IEEE CVPR*, pp. 871–878.

EXTRACTING ORTHOGONAL BUILDING OBJECTS IN URBAN AREAS FROM HIGH RESOLUTION STEREO SATELLITE IMAGE PAIRS

Thomas Krauß^a, Peter Reinartz^a, Uwe Stilla^b

^a German Aerospace Center (DLR), Remote Sensing Technology Institute
PO Box 1116, 82230 Weßling, Germany

^b Technische Universität München, Arcisstr. 21, 80290 München, Germany
thomas.krauss@dlr.de

Commission III, WG III/4

KEY WORDS: VHR data, stereo images, digital surface models, classification, urban modeling, building extraction

ABSTRACT:

Since the number of urban residents is rapidly increasing, especially in developing countries, relatively cheap and fast methods for modeling and mapping such cities are required. Besides the creation and updating of maps from sprawling cities three dimensional models are useful for simulation, monitoring and planning in case of catastrophic events like flooding, tsunamis or earthquakes. With the availability of very high resolution (VHR) stereo satellite data investigations of large urban areas regarding their three-dimensional shape can be performed fast and relatively cheap in comparison to aerial photography – especially for cities in developing countries. Most of the methods actually used for the generation of city models depend on a large amount of interactive work and mostly also on additional information like building footprints and so on. A method for a fully automatic derivation of relatively coarse and simple models of urban structure is therefore of great use. In this paper one approach for such an automatic modeling and a processing chain is sketched and the method used for the modeling of buildings is described.

1 INTRODUCTION

Urban areas all around the world – especially in developing countries – grow rapidly. So more and more 3D models of large city areas are needed for planning and monitoring purposes. For the usage in developing countries such models should be relatively cheap and can be relatively simple. This can be achieved by a fully automatic generation from very high resolution (VHR) satellite image stereo pairs from satellites like Ikonos, QuickBird or the upcoming WorldView series [DigitalGlobe, 2007].

Starting with a stereo scene – a very high resolution satellite image pair – it is possible to generate in a first step a high resolution digital surface model (DSM) by suitable stereo evaluation of the image pair ([Lehner and Gill, 1992], [Krauß et al., 2005], [Hirschmüller, 2005]). Since the ground resolution of the satellites is in the range of one meter the resolution of the surface model is rather coarse in comparison to surface models from airborne camera or lidar data.

Furthermore there exists no additional data in rapidly growing cities. So the proposed automatic process is often limited to only one single stereo image pair. From this only source of information all needed parameters for the generation of a – in a first step coarse – city model have to be extracted.

After the generation of the DSM a digital terrain model (DTM) can be derived from the surface model giving the ground plane (without buildings/trees etc.). Using the high resolution DSM and the satellite images also a true orthophoto can be calculated. Since the satellite data contain four channels (blue, green, red, near infrared) a simple classification based on the orthophoto can be accomplished. This classification uses only a mask of high objects derived from the DSM and DTM and a vegetation

mask calculated from the NDVI of the orthophoto [Krauß et al., 2007].

From this classification all high, non-vegetation objects can be extracted and modeled as described in this paper.

Currently already many approaches exist for city modeling. But these methods are mostly based on cadastral data, aerial images, aerial and terrestrial laser scanner data, terrestrial photographs and more information since the aim of these methods are near photorealistic city models in industrial countries. These models integrate data from several of these sources in often intense manual work for the urban models [CyberCity, 2007, 3D Geo, 2007].

In this paper one part of this processing chain – the automatic extraction of buildings – is described. Many approaches exist which use high resolution airborne lidar data or digital image data. A method often used is the calculation of tensors of inertia and their eigenvectors as main axes of a building. This and a search for maximum diameters of objects and rectangular deviations for describing building outlines as described in [Müller, Zaum, 2005] work only well for convex buildings. A better top-down approach is the recursive rectangle approximation as shown by [Gross, U. et al., 2005]. But in this paper a bottom-up approach for the building outline extraction will be described.

In the approach described here rather simple urban 3D models are generated only from one stereo satellite image pair. Such images are provided at the moment, e.g., by Space Imaging (Ikonos, [SpaceImaging/GeoEye, 2007]) with a ground resolution of about 1 meter panchromatic and 4 m multispectral or in the near future, e.g. by WorldView I (2007) and II (2008), offering half-meter panchromatic and 1.4 to 1.8 m multispectral resolution ([DigitalGlobe, 2007]).

2 DATA

The fully automatic processing chain relies on stereo image pairs which are best acquired in the same orbit with the same illumination conditions, two distinct viewing angles and known internal and external orientation of the satellite (e.g. orbital positions and look angles or RPCs). The processing chain and the building extraction are demonstrated for one Ikonos stereo image pair of the city of Munich.

The scene was acquired on 2005-07-15 at 10:28 GMT with a ground resolution of 83 cm. The viewing angles of the forward and backward image were $+9.25^\circ$ and -4.45° . The images were available only as level 1A product, which are only corrected for sensor orientation and radiometry (0) but contain no further geometric changes.



Figure 1. Section 600 m \times 400 m from the Munich scene (area of Technical University), left and right stereo image

3 PROCESSING CHAIN

The proposed fully automatic processing chain consists of the following steps which are explained shortly in the next sections:

- 1 Preprocessing of the raw imagery
- 2 Creating the digital surface model (DSM)
- 3 Extracting the digital terrain model (DTM)
- 4 Creating true orthophotos
- 5 Classification
- 6 Object extraction
- 7 Object modeling
- 8 Representing the object models through geometric primitives and exporting in suitable 3D format

The part “modeling building objects” of “object modeling” is explained in detail in chapter 4.

3.1 Preprocessing of the raw imagery

In the preprocessing the images are imported and the metadata and the rational polynomial coefficients (RPCs) delivered with the VHR imagery is interpreted. These coefficients are mandatory since they are used to transform the geographical coordinates longitude X , latitude Y and ellipsoid height Z to image coordinates (x,y) by division of two polynoms with 20 coefficients each [Jacobsen et al., 2005, Grodecki et al., 2004]. These geometric calculations are used throughout the generation of the DSM and further for absolute georeferencing and orthophoto generation.

In a further preprocessing step the multispectral channels with much coarser resolution (one pixel multispectral correspondents to four by four pixel in the pan image) than the panchromatic channel are pansharpened.

3.2 Creating the digital surface model (DSM)

In the first processing step a digital surface model is created from the image stereo pair. For implementation of this step in the processing chain some DSM generation methods were evaluated. A classical area based matching approach as described in [Lehner and Gill, 1992] depends on images with few occlusions which means in the case of urban scenes a very narrow viewing angle and so also larger height errors.

More useful for urban scenes seem to be dense stereo algorithms used in computer vision. Such methods depend however on strict epipolar geometry. A good overview of a selection of such algorithms is given on the Stereo Vision Research Page of the Middlebury College maintained by Daniel Scharstein and Richard Szeliski [Scharstein and Szeliski, 2007].

Beside these also two more algorithms based on dynamic programming described in [Krauß et al., 2005, “dynamic line warping”] and [Hirschmüller, 2005, “semi-global matching”] were found applicable for inclusion in the processing chain.

All following investigations were done with results of modified versions of these two dynamic programming algorithms, which becomes necessary due to the non-epipolar geometry of the image pairs. This is due to the fact that the satellite image pair doesn’t allow the creation of a true epipolar image – only so called quasi-epipolar images are possible since the viewing angles of the two satellite images are not parallel.



Figure 2. Digital surface model calculated for a section of 600 m \times 400 m from the Munich scene using the “dynamic line warping” approach

3.3 Extracting the digital terrain model (DTM)

Using this calculated DSM the digital terrain model describing the ground can be derived. This is accomplished by calculating a morphological erosion with a filter size of the maximum of the smallest diameter of all buildings. This results in a height image with every pixel representing the minimum height in this area around the pixel. For calculating the DTM in contrast to [Weidner and Förster, 1995] in reality a median filter returning a rather low order value will be applied instead of the morphological erosion to avoid the domination of the

generated DTM by single outliers from the calculated DSM. After filtering an averaging using the same filter size is applied to obtain a smoother DTM. In the Munich example parts above the DTM are reduced to a flat plane on street level

3.4 Creating true orthophotos

Thanks to the rather dense DSM, the RPCs from the original imagery and the pansharpened multi-spectral stereo images it is possible to derive true orthophotos.



Figure 3. Pan sharpened orthophoto based on the left stereo image and the DSM from the Munich scene

3.5 Classification

Calculating the difference image between DSM and DTM and applying a threshold of “high” (about 4 m) gives the so called “high objects mask” as shown in Figure 4.



Figure 4. High objects mask calculated from the DSM and the derived DTM applying a height threshold of 4 m (section 600 m x 400 m)

The “vegetation mask” is derived from the normalized difference vegetation index (NDVI) which is calculated from the red and near infrared channels of the pansharpened multispectral true orthophoto by applying a suitable vegetation-threshold (Figure 5).

$$NDVI = (NIR - Red) / (NIR + Red)$$

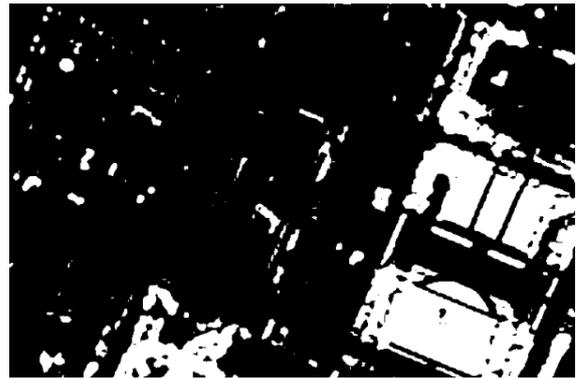


Figure 5. Vegetation mask based on the thresholded NDVI calculated from the orthophoto (600 m x 400 m)

Combining these two binary masks leads to four classes:

- low and no vegetation: streets, plain soil, . . .
- high and no vegetation: buildings, . . .
- low and vegetation: meadows, grass, . . .
- high and vegetation: trees, bushes, . . .

Figure 6 shows these classifications for the used section from the Munich scene:

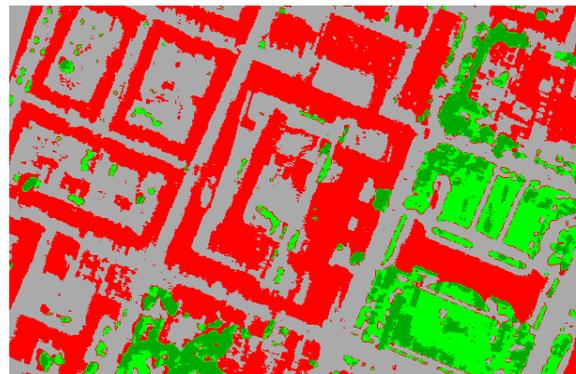


Figure 6. Classification of the Munich scene using a height mask derived from DSM and DTM and a vegetation mask based on the NDVI from the pan-sharpened orthophotos

3.6 Object extraction

For extracting objects the DSM and the orthophoto will be masked with one or more of the derived classes. Extracting the “high vegetation” class yields trees and bushes. The “high non vegetation” class will result mostly in man made buildings. Extracting all “low” objects will result in a ground plane.

3.7 Object modeling

For the simple modeling of the extracted objects following base models are used:

- Model “ground” (class “low”, any type of vegetation)
- Model “tree” (class “high” and “vegetation”)
- Model “building” (class “high” and “no vegetation”)



Figure 7. Simple models used

The “ground” is inserted as a height field extracted from the DTM with an optional texture directly from the true orthophoto.

“Trees” are described by a crown diameter and a treetop height extracted from the classification and the DSM respectively.

“Buildings” are represented as prismatic models restricted to rectangular edges as shown in the next section. In the future the prismatic models will be split to cuboids with optionally parametric roofs.

3.8 Representing the object models through geometric primitives and exporting in suitable 3D format

The coarse models will be represented through geometric primitives. A height field derived from the DTM for “ground” (one for full scene if textured from the true orthophoto), an ellipsoid supported by a cylinder for trees and rectangular vertical walls following the extracted circumference and a horizontal polygonal roof in the first version. A texture may be extracted from the original images by projecting the resulting polygons backward using the RPCs. The optionally textured geometric primitives have to be exported into a suitable 3D vector format. Figure 8 shows the area around the technical university from the Munich scene as shown by a VRML viewer.

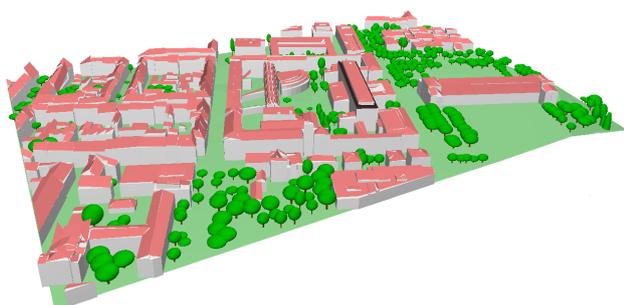


Figure 8. Simple 3D view generated from the Munich scene, size 640 m × 400 m, center: Technical University of Munich, right Old Pinacotheca

4 BUILDING EXTRACTION

For the extraction of buildings the DSM will be masked with one of the objects of type “high, non-vegetation” extracted from the classification. So only one object – the mask enlarged slightly by a morphological dilation – remains in the DSM image.

This (masked) DSM is classified to “height classes” by means of height and optional also by gradients of a small surrounding area of every point. This classification process calculates a height-parameter for each pixel and joins subsequently pixels with height differences below a given threshold together.

Due to this process only clearly by height separable objects remain as classes characterized by the average height of all contributing DSM elements. For example gabled roof will join to one class of half the roof height if the height-join threshold parameter is large enough to join adjacent pixels on the slanted roof together (height-join

parameter of about 1 m, larger as a GSD-step on a roof and smaller than the height of a floor).

In the next step for each of these extracted height classes of the masked object the object outline is extracted. These are shown in green for four selected objects in a section from the Munich test scene in Figure 9.

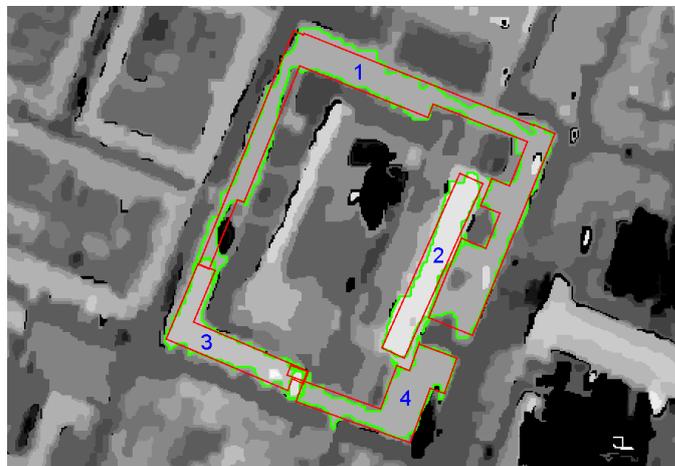


Figure 9. Four selected outlines for building extraction in the Munich scene (full classified DSM of the examined area, not only the one extracted object – image size: 600 m × 400 m)

Based on these outlines of the height class objects a rectangular outline will be calculated. The outline is first parameterized to a vector containing for every pixel the position and an averaged direction between four preceding and four successive points. The angles are combined to full degrees and statistics, showing how many points of the outline possess which direction, is calculated as shown in Figure 10.

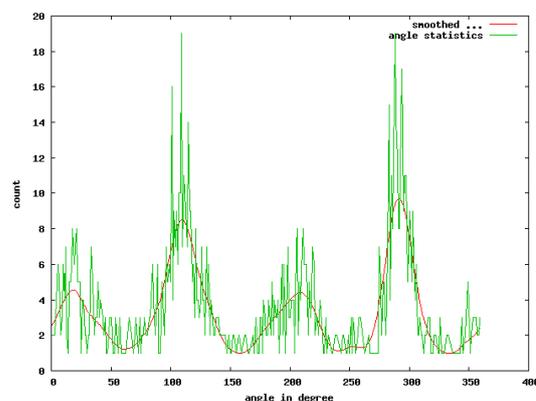


Figure 10. Statistics of measured angles for outline “1” (green outline in Figure 9, right: angle in degree, up: count)

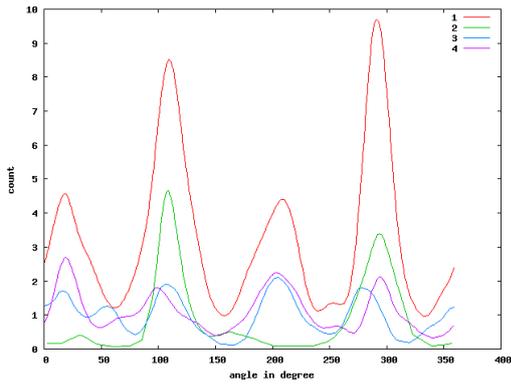


Figure 11. Smoothed distribution for all 4 borders

Smoothing this angular distribution yields Figure 11 and allows the extraction of maxima. These found maxima will be used as the main directions of the outline. For polygon 1 the maxima can be located at 22°, 112°, 202° and 292°.

After filtering and grouping angles together to these found main directions (Figure 12) in a last step consecutive line segments (Figure 13) are intersected and the closed red polygons shown in Figure 9 above are generated.

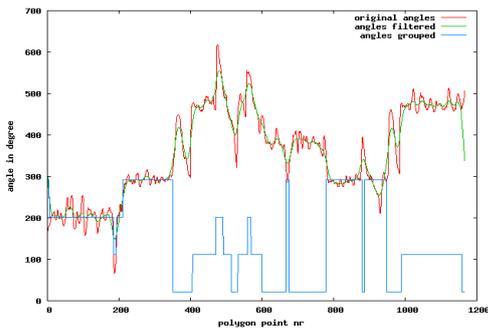


Figure 12. Grouping angles of border elements together to found main directions

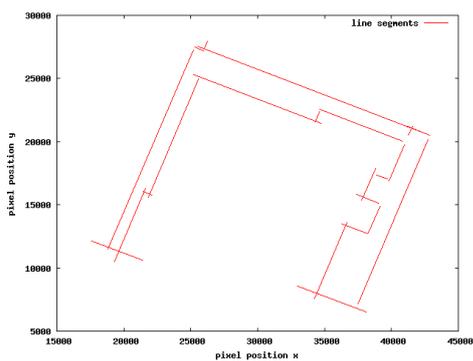


Figure 13. Derived line segments of polygon 1 after grouping of angles

In a first version the coarse modeling is done by simply generating prismatic models with these perpendicular polygons as footprints and flat tops with an average height extracted from the DSM.

Since the accuracy and resolution of the derived DSM is in most cases not sufficient a detailed modeling of the roof shape is difficult. In future versions the polygonal circumference will be divided into rectangles. Using these rectangles an averaged section along and across these rectangles may give a hint if it is a gabled or a flat roof and lead to the selection of a more detailed bottom-up model. In the following example this is evaluated for outline “2” from Figure 9:

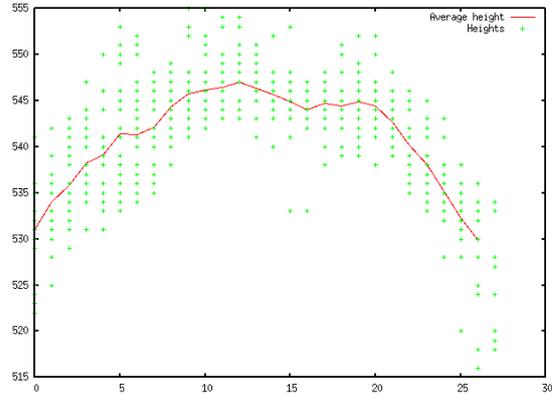


Figure 14. Statistics of DSM heights from object 2 across roof direction (green dots: DSM values all along the roof, red line: averaged profile)

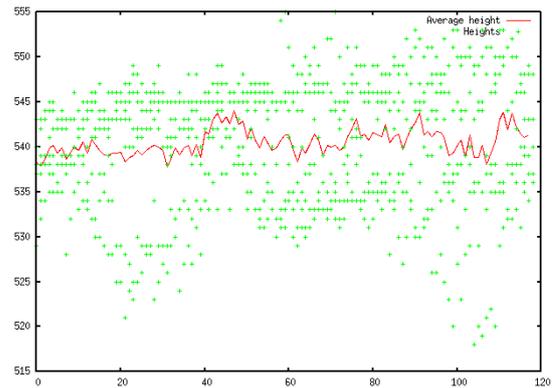


Figure 15. Statistics of DSM-heights from object 2 along roof direction (green dots: DSM values all across the roof, red line: averaged profile)

Since the digital surface model generated from the satellite stereo image pair is in all used generation methods very coarse and noisy, a statistical approach is the only possibility for further estimation of model parameters like the roof type. Also the extraction of smaller roof features like dormers or chimneys are not possible based on the given data.

Since this approach depends on the classification based on height and vegetation it will fail in case of steep rocks not covered by vegetation or buildings covered with roof top gardens or greened roofs.

5 SUMMARY AND OUTLOOK

In this paper a rather simple method for a coarse building extraction and modeling is shown. Also a quick overview of a processing chain for the automatic extraction of three-dimensional city models directly from high-resolution stereo satellite images is given. The processing chain is still in development. So the DSMs generated are not satisfying up to now. Also the automatic extraction of

objects from the classified images exist only in a first evaluation version. The texturing is still missing up to now. But the results gained from each step are encouraging enough to follow the path and refine every step of the chain to receive a new fully automatic system for generating coarse three-dimensional urban models from stereo satellite imagery in a short time.

REFERENCES

- 3D Geo: <http://www.landexplorer.net/> (accessed 07/2007).
- Birchfield, S. and Tomasi, C., 1998: Depth discontinuities by pixel-to-pixel stereo. Proceedings of the 1998 IEEE International Conference on Computer Vision, Bombay, India, pp. 1073–1080.
- CyberCity: <http://www.cybercity.tv/>. (accessed 07/2007).
- DigitalGlobe: <http://www.digitalglobe.com/about/imaging.shtml>. (accessed 07/2007).
- Förstner, W. and Gülch, E., 1987: A fast operator for detection and precise location of distinct points, corners and centres of circular features. In: ISPRS Intercommission Workshop, Interlaken.
- Grodecki, J., Dial, G. and Lutes, J., 2004: Mathematical model for 3D feature extraction from multiple satellite images described by RPCs. In: ASPRS Annual Conference Proceedings, Denver, Colorado.
- Gross H., U. Thoennessen U., v. Hansen W., 2005: 3D-Modeling of Urban Structures. Proceedings of the ISPRS Workshop CMRT 2005, Vienna.
- Hirschmüller, H., 2005: Accurate and efficient stereo processing by semi-global matching and mutual information. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
- Jacobsen, K., Büyüksalih, G. and Topan, H., 2005: Geometric models for the orientation of high resolution optical satellite sensors. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 36 (1/W3). ISPRS Workshop, Hannover.
- Krauß, T., Reinartz, P., Lehner, M., Schroeder, M. and Stilla, U., 2005: DEM generation from very high resolution stereo satellite data in urban areas using dynamic programming. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 36 (1/W3). ISPRS Workshop, Hannover.
- Krauß, T., Reinartz, P. Lehner, M., 2007: Modeling of urban areas from high resolution stereo satellite images. In: Proceedings of ISPRS Workshop “High-Resolution Earth Imaging for Geospatial Information”, Mai 29 – Juni 1, 2007, Hannover.
- Lehner, M. and Gill, R., 1992: Semi-automatic derivation of digital elevation models from stereoscopic 3-line scanner data. ISPRS, 29 (B4), pp. 68–75.
- Müller S., Zaum D., 2005: Robust Building Detection in Aerial Images. Proceedings of the ISPRS Workshop CMRT 2005, Vienna.
- Otto, G. and Chau, T., 1989: Region growing algorithm for matching of terrain images. Image and vision computing (7) 2, pp. 83–94.
- Scharstein, D. and Szeliski, R. Middlebury stereo vision page: <http://cat.middlebury.edu/stereo>. (accessed 07/2007).
- Weidner, U., Förstner, W., 1995: Towards automatic building extraction from high resolution digital elevation medels. ISPRS J. 50 (4), 38–49.
- SpaceImaging/GeoEye: <http://www.geoeye.com/> (accessed 07/2007).

AUTOMATIC TRAFFIC MONITORING WITH AN AIRBORNE WIDE-ANGLE DIGITAL CAMERA SYSTEM FOR ESTIMATION OF TRAVEL TIMES

F. Kurz^{a*}, B. Charmette^a, S. Suri^a, D. Rosenbaum^a, M. Spangler^b, A. Leonhardt^b,
M. Bachleitner^c, R. Stätter^a, P. Reinartz^a

^aGerman Aerospace Center (DLR), Remote Sensing Technology Institute, PO Box 1116, D-82230 Weßling, Germany

^bMunich University of Technology (TUM), Traffic Engineering and Control, Arcisstr. 21, D-80290 München, Germany

^cGeneral German Automobile Association (ADAC), Development of traffic information, Am Westpark 8, D-81373 München, Germany
franz.kurz@dlr.de

Commission 3 WG III/5

KEY WORDS: Aerial cameras, image series, traffic parameters, road detection, vehicle detection

ABSTRACT:

Knowledge of accurate travel times between various origins and destinations is a valuable information for daily commuters as well as for security related organizations (BOS) during emergencies, disasters, or big events. In this paper, we present a method for automatic estimation of travel times based on image series acquired from the recently developed optical wide angle frame sensor system (3K = "3-Kopf"), which consists of three non-metric off-the-shelf cameras (Canon EOS 1Ds Mark II, 16 MPixel). For the calculation of overall travel times, we sum up averaged travel times derived from individual vehicle velocities to pass defined road segments. The vehicle velocities are derived from vehicle positions in two consecutive geocoded images by calculating its distance covered over time elapsed. In this context, we present an automatic image analysis method to derive vehicle positions and vehicle distances involving knowledge based road detection algorithm followed by vehicle detection and vehicle tracking algorithms. For road detection, we combine an edge detector based on Deriche filters with information from a road database. The extracted edges combined with the road database information have been used for road surface masking. Within these masked segments, we extract vehicle edges to obtain small vehicle shapes and we select those lying on the road. For the vehicle tracking, we consider the detected vehicle positions and the movement direction from the road database which leads to many possible matching pairs on consecutive images. To find correct vehicle pairs, a matching in the frequency domain (phase correlation) is used and those pairs with the highest correlation are accepted. For the validation of the proposed methods, a flight and ground truth campaign along a 16 km motorway segment in the south of Munich was conducted in September 2006 during rush hour.

1. INTRODUCTION

Near real time monitoring of natural disasters, mass events, and large traffic disasters with airborne SAR and optical sensors will be the focus of several projects in research and development at the German Aerospace Center (DLR) in the next years. In this overall frame, knowledge of accurate travel times between various origins and destinations is a valuable information not only for traffic management and information purposes but also for security related organizations (BOS) during emergencies, disasters, or big events. One application currently under development is the generation of isochronal maps derived from airborne imagery, which shows the up-to-date travel times from each map position to the accident scene. This may support BOS coordinators for fleet disposition and adequate alerting in times of adversities.

In general, travel time estimation is based on data from conventional stationary measurement systems such as inductive loops, radar sensors or terrestrial cameras. One handicap of these methods is the low spatial resolution depending on the ground distribution. New approaches include data by means of mobile measurement units which flow with the traffic (floating car data, FCD, (Schaefer et al., 2002), (Busch et al. 2004)).

The big advantage of the remote sensing techniques presented here is that the measurements can be applied nearly everywhere (exception: tunnel segments) and do not depend on any third party infrastructure. Besides, airborne imagery provides a high spatial resolution combined with acceptable temporal resolution depending on the flight repetition rate, but require complex

image analysis methods and traffic models to derive the desired traffic parameters (Ernst, 2005).

In this paper, we present a method for automatic estimation of travel times based on geocoded image series acquired from a recently developed DLR wide angle frame sensor system. For the calculation of the point-to-point travel times, we apply a simple semi-empirical travel time method which is based on averaged travel times derived from individual vehicle velocities to pass defined road segments. For this, individual vehicle positions and velocities are required which are provided by automatic image analysis tools.

In this context, we present an automatic image analysis method to derive vehicle positions and vehicle distance covered based on a road detection algorithm using a road database followed by vehicle detection, and vehicle tracking algorithms.

2. ESTIMATION OF TRAVEL TIMES

While only few detection methods offer the possibility to directly measure travel times of single cars (such as vehicle re-identification by licence plate recognition, floating car tracking), certain model assumptions are needed to derive travel times from local detector data.

Different methods for the estimation of travel times are applied depending on the sensors used and on the type and availability of sensor data. Simple methods derive travel times by constructing pseudo vehicle trajectories from time mean speed measured locally. Those methods are easy to implement and need no calibration but have difficulties to capture traffic

dynamics such as traffic jams at high densities. Better results can be generated by applying traffic flow models that are widely used in traffic engineering and control. For simulation or traffic prediction, data from inductive loops, terrestrial cameras, variable speed limits and other context data are fed into these traffic models.

Macroscopic traffic models like METANET (Kotsialos, 2002) reproduce the spatial and temporal traffic flow based on averaged microscopic car dynamics and are used to estimate or predict traffic state information. A large group of macroscopic models is based on the fluid-dynamic theory, interpreting the traffic flow as a compressible fluid (e.g. Daganzo, 1997).

Microscopic traffic models (e.g. Krauß, 1997) describe the vehicles with detailed behaviour. Each vehicle's reactions to the actual traffic situations are modelled by assumptions concerning car following (e.g. Gipps, 1981) and overtaking procedures (e.g. Gipps, 1986). Microscopic models are more detailed than macroscopic models, but they require significantly higher computational efforts, what limits their use for real-time applications. A class of microscopic traffic models based cellular automat technology tries to overcome those problems (e.g. Nagel, 1992).

New approaches aim to fuse all relevant data sources such as floating car data, stationary sensor data, modelled data and historic information..

In case of airborne optical data, travel times for a platoon of vehicles can be recorded from a helicopter, rather than a single car. The helicopter simply follows the vehicles in a traffic jam (Angel, 2003).

In case of wide-angle airborne data, like the DLR 3K camera, we propose an instantaneous method to derive overall travel times T for a road section. For this, we divide the road section in N equally spaced segments with length L and calculate mean vehicle velocities \bar{v}_i for segment i . The overall travel time T is then the sum of travel times for each segment.

$$T = \sum_{i=1}^N \min \left\{ \frac{L}{\bar{v}_i}, \frac{L}{v_{\min}} \right\} \quad (1)$$

In case of very low mean vehicle velocities, the travel times tend to be very high (standing cars would lead to an infinite value). Thus, a minimum vehicle velocity v_{\min} is introduced to avoid too high travel times. The minimum vehicle velocity parameter is empirical and must be adjusted to the traffic type as well as to the road segment monitored as e.g. in case of blocked motorway lanes the estimated travel times will be erroneous using a constant minimum vehicle velocity for all traffic types.

In this paper, we focus on congested traffic with traffic densities higher than the critical density without additional disturbances like blocked lanes and we set *a priori* the minimum vehicle velocity to 7.2 km/h following the METANET settings.

The mean vehicle velocities \bar{v}_i will be derived from the 3K image sequences based on the vehicle positions in at least two consecutive geocoded images and based on the time elapsed (Hinz, 2007).

In the following chapter, automatic image analysis tools to derive vehicle positions and velocities will be presented.

3. AUTOMATIC ROAD AND VEHICLE DETECTION

For an effective real time traffic analysis, the road surface needs to be clearly demarcated. Thus, we automatically delineate the roadsides by two linear features using the processing chain illustrated in Fig 1. Automated road extraction has been

developed as an independent module in our ongoing research on real time traffic analysis. For brevity, we are not explaining the road extraction procedure in details. The road extraction starts by forming a buffer zone around the roads surfaces using a road database as basis for the buffer formation process. In the marked buffer zone, we use edge detection and feature extraction techniques. The critical step of edge detection is based on an edge detector proposed by Phillippe Paillou for noisy SAR images (Paillou, 1997). Derived from Deriche filter (Deriche, 1989) and proposed for noisy SAR images, we found this edge detector after ISEF filtering (Shen and Caston, 1992) extremely efficient for our purpose of finding edges along the roadsides and suppressing any other kind of surplus edges and noise present. The roadside identification module, again with the help of the road database tries to correct possible errors (gaps and bumps) that might have creped in during the feature extraction phase.

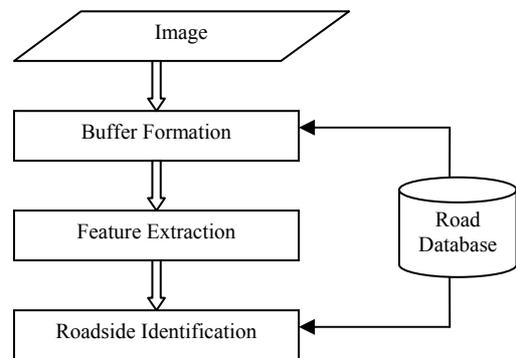


Fig 1 Implemented processing chain for a knowledge based road extraction

With the information of the roadside obtained in the processing step described before, it is possible to perform vehicle detections and tracking limited only to the roads. For this, we developed an algorithm for the detection and tracking of vehicles which is described in the following (see Fig 2).

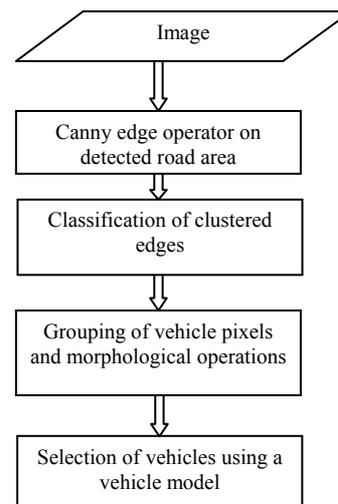


Fig 2 Vehicle detection algorithm

Based on information about the alignment and direction of the roadside, all pixels of the road including the road direction are marked. For the vehicle detection, a Canny edge operator is

applied and a histogram on the edge steepness is calculated. Then, a k-means algorithm is used to split edge steepness statistics into three parts which represent the three main classes, vehicles, roads, and not classifiable.

We consider the part with the lowest steepness being mainly pixels of the road background, since its intensity is quite uniform. Besides, we assume that the part with the highest steepness is due to the high discontinuity in the intensity probably populated by vehicles. This part of the statistic is also contaminated by the paintings on the roadside, shadows, sign boards, trees, etc.

In the part which is not classifiable, it must be determined which pixels belong to the road background or to potential vehicles. For the decision, the pixel neighbourhood is examined. Pixels directly connected with a potential vehicle pixel are moved into the vehicle class. Remaining pixels are finally considered as road background.

In case of coloured images, the processes described above are realized on each channel separately in order to obtain a multi channel edge image.

In the next step, the roadside pixels are eliminated from the part with higher steepness. As the roads are well determined by the road extraction, the roadside lines can be found easily. Thus, the algorithm erases all pixels with high edge steepness which are laying on a roadside position. Thereby, it avoids erasing vehicles on the roadside by considering the width of the shape. Mid-line markings are erased using a dynamic threshold detector. This is done in order to reduce false detections, since these mid-line markings may look like white cars.

Then, potential vehicle pixels are grouped by selecting neighbored pixels. Each group is considered to be composed of potential vehicle pixels connected to each other. With the groups obtained a list of potential vehicles is produced.

In order to extract real vehicles from the potential vehicle list, a closing of the shapes of the potential vehicles is performed. We can see the effect of the closing algorithm in Fig 3.

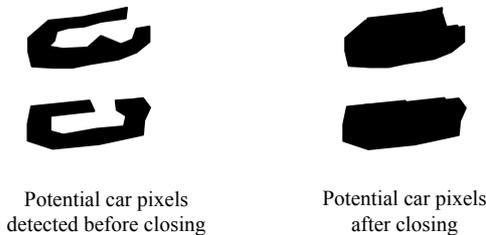


Fig 3. Closing the shapes of potential car pixels

Using the closed shape, the properties of vehicle shapes are described by its direction, its area, the length and width following the direction, and its position on the road. Based on these parameters, the vehicles are assumed to have rectangular shapes with a specific length and width oriented in the road direction. Their area should be about the length multiplied by the width and vehicles must be located on the roads. We set the values for the vehicle length to 5.7 m and for the width to 2.6 m (for standard cars). In case of detections with very low distances the algorithm assumes a detection of two shapes for the same vehicle. Then, it merges the two detections into one vehicle by calculating averages of the positions.

Finally, based on this vehicle model, a quality factor for each potential vehicle is found and the best vehicles are chosen.

By applying vehicle detections to an image sequence taken within a small time interval, which is the case for the 3K image series, we are able to match a vehicle in this sequence, in order to measure its velocity.

Within the tracking algorithm (see Fig 4), a search area for each vehicle detected in the first image is defined in the second image based on the predicted position. During short time steps between two images, vehicles do not change their direction significantly, so it can be assumed that the matching vehicle from the other image lies within the search area. Thus, we can find potential corresponding pairs of detected vehicles. However, in case of a high vehicle density on the images, as it is the situation in traffic congestions, many vehicles in the second image are located in the search area of one vehicle in the first image. At the same time each car in the second image has lots of possible origin cars in the first image. In order to choose the right pair, it is checked that the direction of both vehicles in a possible vehicle pair is the same. As criteria, the road direction from the road database is considered for the selection correct pairs.

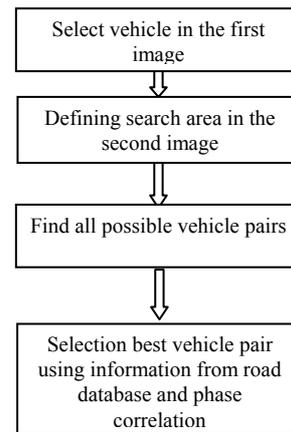


Fig 4 Car tracking algorithm

Then, a phase correlation algorithm to find best corresponding vehicles is applied for each pair by using small image patches around the car. With these tests a quality coefficient is determined for each pair of matching candidates. At the end, when a car in the first image has several correspondents in the second image, we keep only the pair with the best quality value.

4. SENSORS AND DATABASE

4.1 3K camera system

The 3K camera system (3K = “3Kopf”) consists of three non-metric off-the-shelf cameras (Canon EOS 1Ds Mark II, 16 MPix). The cameras are arranged in a mount with one camera looking in nadir direction and two in oblique sideward direction (Fig 5), which leads to an increased FOV of max 110°/31° in across track/flight direction.



Fig 5. DLR 3K-camera system consisting of three Canon EOS 1Ds Mark II, integrated in a ZEISS aerial camera mount

The camera system is coupled to a GPS/IMU navigation system, which enables the direct georeferencing of the 3K optical images. Fig 6 illustrated the image acquisition geometry of the DLR 3K-camera system. Based on the use of 50 mm Canon lenses, the relation between airplane flight height, ground coverage, and pixel size is shown, e.g. the pixel size at a flight height of 1000 m above ground is 15 cm and the image array covers up 2.8km in width.

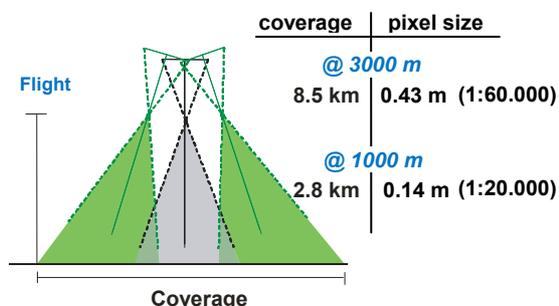


Fig 6. Illustration of the image acquisition geometry. The tilt angle of the sideward looking cameras is approx. 35°.

4.2 Test-Site

The motorway A8 south of Munich is one of the busiest parts of the German motorway network with an average load of around 100.000 vehicles per day. Fig 7 shows the 16 km motorway section between motorway junctions “Hofolding” and “Weyarn”, which was selected as test site on 2. Sep. 2006. At this time, heavy traffic was expected at this section caused by homebound travellers.

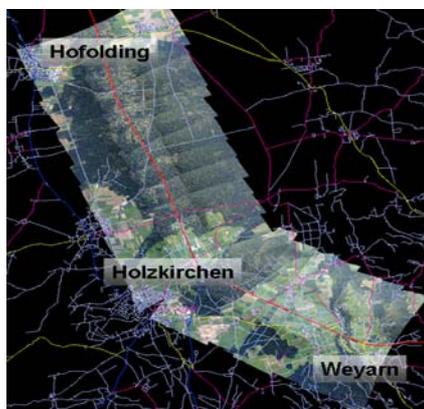


Fig 7. 16km motorway strip (A8) south of Munich as imaged by 3K camera system

4.3 3K imagery

Three 3K data takes were acquired on 2. Sep. 2006 between 14:01 and 15:11 from 2000m above ground. Table 1 lists the exact acquisition times of each data take. During each overflight, 22 image bursts were acquired each containing four consecutive images. The time difference within these bursts was 0.7 s, so that each car was monitored at least 2.1 s.

ID	Date	Pixel size	Images	H.a.G.
3K-1a	02-Sep-2006 14:01-14:12	30cm	22x4x3	2000m
3K-1b	02-Sep-2006 14:30-14:40	30cm	22x4x3	2000m
3K-1c	02-Sep-2006 15:01-15:11	30cm	22x4x3	2000m

Table 1 3K camera data takes at A8 south of Munich

For further analysis, 3K images were geocoded using onboard GPS/IMU measurements with an absolute position error of 3m in nadir images and less than one pixel relative. The last error has great influence on the derived vehicle velocities.

4.4 Road database and ancillary data

Data from a road database are used as *a priori* information for the automatic detection of road area and vehicles. One of these road databases has been produced by the NAVTEQ Company (NAVTEQ 2006). The roads are given by polygons which consist of piecewise linear “edges,” grouped as “lines” if the attributes of connected edges are identical. Up to 204 attributes are assigned to each polygon, including the driving direction on motorways, which is important for automated tracking. Recent validations of position accuracy of NAVTEQ road lines resulted in 5m accuracies for motorways.

Data from other sources and sensors were collected to make an overall comparison of derived travel times. Although traffic messages from the traffic channel do not contain travel times but congestion lengths, this information represents the state of the art and was therefore collected. Additionally, data from local detectors are used to calculate instantaneous travel times:

$$TT(t) = \sum_{i=1}^n \frac{l_i}{v_i(t)} \quad (2)$$

With $TT(t)$ being the travel time for the whole stretch at time instant t , v_i being the speed reported at detector station i and l_i being the length of the segment assigned to detector station i .

ID	Time	Traffic message channel	Travel times from detector data
3K-1a	-14:06	Halting traffic 14km	34 min
	-14:11	Congestion 7km	
3K-1b	-14:40	Congestion 12km	28 min
	-15:03	Halting traffic 18km	
3K-1c	-15:07	Congestion 7km	26 min
	-15:11	Congestion 12km	

Table 2 Traffic message channel information and estimated travel times from detector data for A8 south of Munich

Table 2 lists data from other sources and sensors, which are linked to the acquired 3K data takes. Obviously, the traffic messages vary strongly within short time periods between congested and halting traffic.

4.5 Reference vehicle

As ground truth for travel times, two runs with a GPS equipped vehicle (ADAC) in northbound direction and one run in southbound direction was conducted. Table 3 lists the links between reference vehicle runs to 3K data takes. It should be mentioned, that a direct comparison between travel times from reference vehicle and 3K data takes contain systematic errors, as the northbound runs take around half an hour and the 3K data set represent a time span of ten minutes.

ID	Date	Strip	Travel-time	Data take	
R1	02-Sep-2006 14:02-14:36	16km	35 min	3K-1a	North-bound
R2	02-Sep-2006 14:38-14:47	16km	9 min	3K-1b	South-bound
R3	02-Sep-2006 14:48-15:18	16km	31 min	3K-1c	North-bound

Table 3 Data from reference vehicle at A8 south of Munich

4.6 Manual measurements

For manual measurements, the 3K images have been mosaicked in two orthophotos according to Fig 7 (taking the 2. and 4. image of the bursts). This mosaic represents the traffic situation with a relative time difference of 1.4s. Measurement of vehicle positions and corresponding vehicles resulted in the velocity of vehicles. These measurements were basis for calculating the travel times which are illustrated in Table 4.

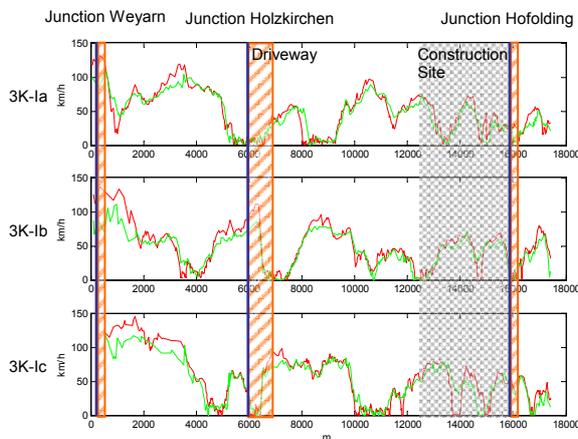


Fig 8. Manual measured northbound velocity profiles on the left lane (red) and right lane (green) for all 3K datatakes

Fig 8 shows velocity profiles of all vehicles going northbound during the three 3K data takes. The vehicle velocities vary strongly between 150 km/h and 0 km/h with small differences between left and right lane.

5. RESULTS

5.1 Comparison of travel times

Travel times are calculated using measured vehicle positions and velocities based on equation (1). For this, the minimum vehicle velocity was set to 7.2 km/h according to the METANET settings, and average vehicle velocities \bar{v}_i were calculated based on segments L of 1 km length. Table 4 lists the travel times derived from the 3K data sets separated in left and right lane, the travel times from the detector data, and the travel times of the reference vehicle.

In general, left lane travel times are slightly shorter than right lane travel times. In comparison with the reference travel times, the 3K derived times are higher and the detector derived times are shorter. Besides, the 3K times decrease with higher minimum vehicle velocity or longer road segments. Thus, with a minimum vehicle velocity of 15 km/h or a road segment length of around 10 km, the derived travel times correspond with the reference times. These two parameters are free and

must be adjusted to the road type and traffic type, but in this experiment they were fixed to the METANET settings.

ID	Ref	Travel time 3K		Detector data	
		Left lane	Right lane		
1a	35'	39'48''	39'57''	34'	S-N
1b	X	38'36''	40'42''	28'	S-N
1b	9'	08'04''	X	X	N-S
1c	31'	37'06''	37'27''	26'	S-N

Table 4 Comparison of travel times derived from 3K images with reference travel times

Instantaneous travel times derived from detector data do not completely comply with the reference measurements. That is because only speeds at certain stations are used that are more or less a set of random observations of the complete freeway stretch. Also, instantaneous speeds do not reflect the dynamics of traffic, as they do not consider the time a vehicle actually needs to pass the different stations.

5.2 Automatic detection of vehicles



Fig 9. Zoom into the two images used for tests on the automatic vehicle detection. Rectangles mark automatic detections, triangles point into the travel direction.

We tested our programs for automatic road and vehicle detection on several image sequences of traffic on motorways taken by the 3K camera system. Fig 9 shows a zoom into two example images which were taken from a sequence obtained at

the motorway A8 near "Holzkirchen". The lower image was taken 1.4s later than the upper image. Using these images the programs for automatic road and vehicle detection were tested (see also section 3), and the results of the automatic detection were compared to manual measured reference.

Table 5 shows the results of the automatic and manual detection of cars on these two images as well as the completeness and correctness of the automatic car detection with respect to the reference. It can be seen that the correctness is quite high, whereas the completeness is lower. We used these results of the automatic vehicle detection as input for the vehicle tracking program. However, the results in correctness and completeness of corresponding pairs of vehicles in these two images obtained from the vehicle tracking program are quite low due to the low completeness of the automatic vehicle detection.

ID	Total automatic detections	Correct automatic detections	Manual detections	Correctness of automatic detection	Completeness of automatic detection
Ia	117	95	193	81%	49%
Ib	56	47	170	84%	28%

Table 5 Comparison of automatic and manual car detection for two images

6. CONCLUSIONS

The investigations show the high potential to use airborne image time series for the estimation of travel times. Firstly it is shown that the automatic road detection works sufficiently good for restricting the search area of the vehicle detection and for giving information about road directions. Also the automatic vehicle detection has already reached a high level of accuracy concerning the correctness, although the completeness is still too low for car tracking applications and has to be improved. A further challenge of the methodology is a better estimation of the minimum velocity for modelling the vehicle speed in a traffic congestion. Models have to be established to estimate this parameter out of vehicle density and/or other measurable values. It is expected that overall accuracy could be enhanced if the system would be combined with a dedicated traffic flow model in order to utilize the strength of both approaches.

The results of such measurements can be used for obtaining travel times and other relevant traffic parameters in cases of catastrophes or other special events (e.g. mass events), where the costs for an airborne data acquisition is justified.

It is planned to test the airborne based camera travel time measurement system against a real reference database with high sampling rate, e.g. vehicle re-identification via automatic licence plate recognition.

7. REFERENCES

- Angel, A., Hickman, M., Mirchandani, P., Chandnani, D. (2003). *Methods of Analyzing Traffic Imagery Collected From Aerial Platforms*. IEEE Transactions on Intelligent Transportation Systems, Vol 4 No 2 pp 99 – 107
- Busch, F.; Glas, F.; Bermann, E. (2004). *Dispositionssysteme als FCD-Quellen für eine verbesserte Verkehrs-lagerekonstruktion in Städten - eine Überblick*. Straßenverkehrstechnik 09/04
- Daganzo, F. C. (1997). *Fundamentals of Transportation and Traffic Operations*. Elsevier Science Inc., New York.
- Deriche, R. (1987). *Using Canny's criteria to derive an optimal edge detector recursively implemented*. The International Journal on Computer Vision, Vol. 1, No. 2, pp 167-187
- Ernst, I., Hetscher, M., Zuev, S., Thiessenhusen, K.U., Ruhé, M. (2005). *New approaches for real time traffic data acquisition with airborne systems*. In: TRB Transportation Research Board, TRB 2005, Washington.
- Gipps, P.G. (1981). *A behavioural car-following model for computer simulation*. Transportation Research Part B, 20, 403-414.
- Gipps, P.G. (1986). *A model for the structure of lane-changing decisions*. Transportation Research Part B, 15, 105-111.
- Hinz, S., Kurz, F., Weihing, D., Meyer, F., Bamler, R. (2007). *Evaluation of Traffic Monitoring based on Spatio-Temporal Co-Registration of SAR Data and Optical Image Sequences*. PFG – Photogrammetry – Fernerkundung – Geoinformation, in print.
- Kotsialos, A., Papegeorgiou M., Diakaki, C., Pavlis, Y., Moddelham, F. (2002). *Traffic Flow Modeling of Large-Scale Motorway Networks Using the Macroscopic Modeling Tool METANET*. IEEE Transactions on Intelligent Transportation Systems. Vol 3 No 4, pp 282–292
- Krauß, S., Wagner, P., Gawron, C. (1997). *Metastable states in a microscopic model of traffic flow*. Physical Review E, 5, pp 5597 - 5602
- Kurz, F., Müller, R., Stephani, M., Reinartz, P., Schroeder, M. (2007). *Calibration of a wide-angle digital camera system for near real time scenarios*. In: Heipke, C.; Jacobsen, K.; Gerke, M. [Hrsg.]: ISPRS Hannover Workshop 2007, High Resolution Earth Imaging for Geospatial Information, Hannover, 2007-05-29 - 2007-06-01, ISSN 1682-1777
- Nagel, K.; Schreckenberg, M. (1992). *A cellular automaton model for freeway traffic*. J. Phys. I France 2 (1992) pp 2221-2229.
- Paillau, P. (1997). *Detecting Step Edges in Noisy SAR Images: A New Linear Operator*. IEEE Transactions on Geoscience and Remote Sensing, Vol. 35, No.1, pp 191-196
- Schaefer, R.-P., Thiessenhusen, K.-U., Wagner, P. (2002). *A traffic information system by means of real-time floating-car data*. Proceedings of ITS World Congress, October 2002, Chicago, USA.
- Shen, J., and Castan, S. (1992). *An optimal linear operator for step edge detection*. CVGIP, Graphics Models and Image Processing, Vol. 54, No. 2, pp 112-133

INFERRING TRAFFIC ACTIVITY FROM OPTICAL SATELLITE IMAGES

J. Leitloff^{a,*}, S. Hinz^b, U. Stilla^a

^a Photogrammetry and Remote Sensing

^b Remote Sensing Technology

Technische Universität München, Arcisstrasse 21, 80333 München, Germany

{Jens.Leitloff | Stefan.Hinz | Uwe.Stilla}@bv.tum.de

KEY WORDS: Urban, Detection, Matching, Quickbird, Multispectral, Multiresolution

ABSTRACT:

In this paper we describe an approach to automatically estimate movements of vehicles in optical satellite imagery. The approach takes advantage of the fact that the optical axes of the panchromatic and multispectral channels of current spaceborne systems like IKONOS or Quickbird are not coinciding. The time gap that appears between the acquisition of the panchromatic and multispectral data can be used to derive velocity information. We employ a sub-pixel matching approach relying on gradient directions followed by least-squares fitting of Gaussian kernels to estimate the movement. The incorporation of the least-squares framework provides the basis to conclude about the accuracy of the movement estimates and to apply a statistical test deciding whether an object moves at all. We illustrate the matching and estimation scheme by various examples of real data.

1. INTRODUCTION

The automatic detection, characterization and monitoring of traffic using airborne and spaceborne data has become an emerging field of research. Approaches for vehicle detection and monitoring include not only video cameras but nearly the whole range of available sensors such as optical aerial and satellite sensors, infrared cameras, SAR systems, and airborne LIDAR. The broad variety of approaches can be viewed, for instance, in the compilations (Stilla et al., 2005) and (Hinz et al., 2006). Although airborne cameras are already in use and seem to be an obvious choice, satellite systems have entered the resolution regime required for vehicle detection. Sub-metric resolution is available in the optical domain and, since the successful launch of TerraSAR-X, closely followed by Spotlight SAR data. While the utilization of along-track interferometric spaceborne SAR data as delivered by TerraSAR-X is straightforward for movement estimation, optical satellite images do not provide interferometric capabilities nor they allow for acquiring image sequences with reasonable frame rate (e.g., the time gap of IKONOS single-pass stereo pairs reaches 6-12s).

However, the CCD linescanners implemented in the IKONOS and Quickbird instrument offer the potential for deriving information about moving objects. Due to constructional reasons the panchromatic (pan) and multispectral (ms) channels have non-coinciding optical axes, which lead to a small time gap of 0.2s between the acquisition of the pan and ms images (see e.g. (EURIMAGE, 2007)).

Two typical traffic scenes of a pan-sharpened Quickbird scene of an urban area are shown in Figure 1. The delay of 0.2s can be clearly seen by the typical color fringe caused by fast moving objects, which are observed at different positions in the pan and ms data.

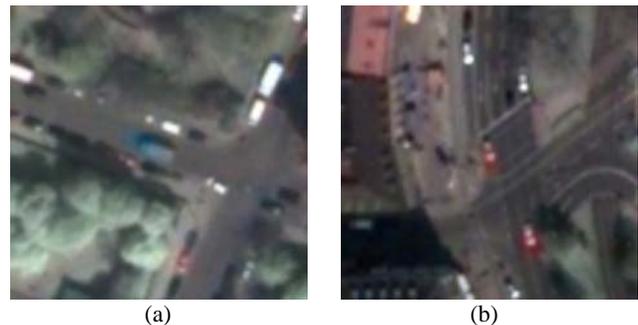


Figure 1. Spaceborne traffic scenes (Quickbird). Moving cars can be identified by their color fringe; see blue car in (a) and red cars in (b)

The appearance of color fringes at moving objects is well-known and has been mostly treated as artefact. Usually techniques are developed for removing the fringes to enhance the visual quality of images, although the potential for movement estimation has been already recognized – especially for airborne line scanner cameras like HRSC or ADS-40. In the context of spaceborne images, for instance, Etaya et al. (2004) showed the potential for moving object detection, using the delay between the acquisition of the pan and the ms channels. While this work was also done on Quickbird imagery another successful application of their approach is demonstrated for Spot data in Etaya et al. (2005). Here ocean wave movements during the Northern Sumatra Earthquake in 2004 were detected.

While these works use manual object selection in both channels, we focus on the automatic matching of hypotheses found in the high resolution channel with their conjugates in the lower resolution channels. In particular, we apply the approach to the task of automatic movement detection of vehicles.

* Corresponding author.

2. MOVEMENT ESTIMATION

First, an overview of the complete system for vehicle extraction will be given. Afterwards the procedures for movement estimation are described in detail.

2.1 System Overview

The presented work is a part of an overall framework for the estimation of traffic parameters from satellite imagery. As input the panchromatic and multispectral channels are used. These images are geo-coded and co-registered with road data taken from a geoinformation system (GIS). These give information about potential regions of interest, i.e. roads, junctions, and parking lots. The further processing is divided into the extraction of vehicles that are grouped in queues or rows and the detection of isolated standing vehicles. The vehicle queue detection has already been described in detail, see (Leitloff et al., 2006), and approaches for the extraction of single cars from sub-meter resolution images are outlined in (Hinz, 2005, Hinz et al., 2007).

This work focuses on the movement estimation of single vehicles. To thoroughly analyze the robustness of the approach, we assume the input data to be 100% complete and correct. Therefore, vehicles were selected manually from the panchromatic channel and are used as input. This step can easily be replaced by the automatic techniques mentioned above. A simplified system overview is illustrated in Fig. 2.

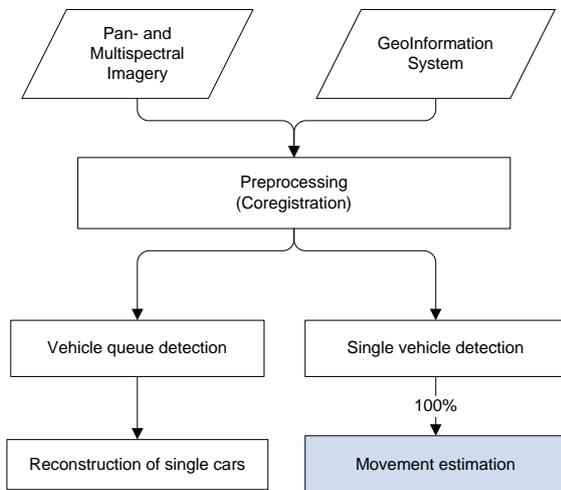


Figure 2. System overview

2.2 Initial hypotheses extraction

Starting from the manually selected vehicles' center in the pan image the position of the corresponding blob in the low resolution ms images needs to be determined. Instead of applying a search algorithm on all color channels independently, the RGB images are transformed into the Intensity-Hue-Saturation (IHS) color space. The distribution of hue and saturation values for different typical objects is illustrated in Figure 3. It depicts a horizontal slice through the IHS conic, with the intensity values projected on the slice, i.e. a polar coordinate system is used, where the length of the vector

between each point and the center corresponds to the saturation and the hue is defined as angle between abscissa and the vector to each point. It comes clear that high color saturation of an object like a red or blue car on a road in at least one of the multispectral channels strongly influences the distance from the coordinate origin (Figs. 3a,b), while a gray road with less saturation is located near the origin (Fig. 3c).

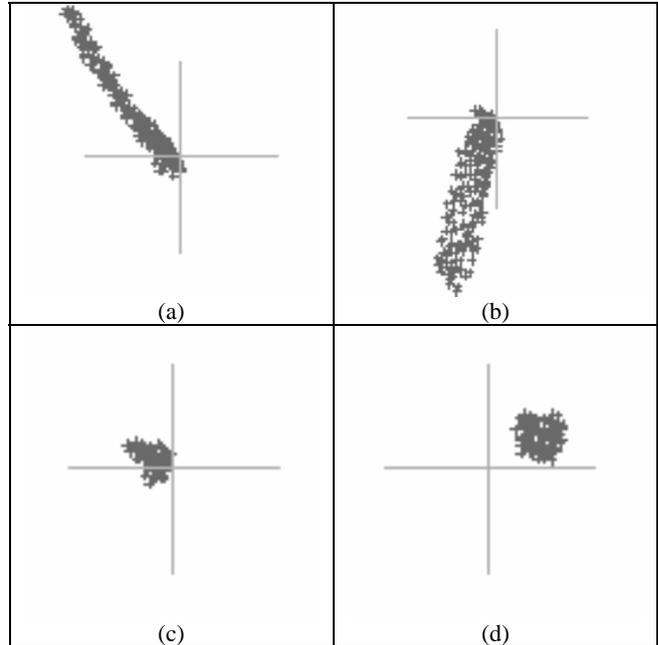


Figure 3. Hue (angle) and saturation (distance) for (a) red car and street, (b) blue car and street, (c) street, (d) grassland

Examples for the saturation channel are illustrated in Figure 4. The blob structure of vehicles in this channel becomes clearly visible and is fundamental for the following steps.

To find the translation vector between the blob of the input region and its conjugate region in the ms data, a sub-pixel matching approach is employed. The approach relies on the work of Steger (2001). The main idea of this algorithm is the use of a gradient filter to determine the gradient direction for each pixel of an image. First, a template is cropped from the pan channel around the vehicles' position. To adapt it to the scale of the coarser resolution of the ms channels the template is smoothed by a Gaussian kernel for eliminating possible substructures from the pan image. Then the gradient directions for this template and the IHS channels are calculated. An exhaustive search constrained to a reasonably limited area compares the gradient directions of the template image with the gradient directions found in the IHS images, which results in a similarity measure for each pixel (Figure 5a and 5b). Notice that the pure utilization of gradient directions makes normalization or contrast manipulations of one or both data sets virtually unnecessary. For eventually detecting the vehicle region in the ms data, the position showing the highest similarity is chosen, if a maximum distance to the position in the pan channel is not exceeded. Figure 5c and 5d show the results of the initial detection.



Figure 4. (a) and (c) Pansharpened images, (b) and (d) corresponding saturation images

2.3 Accurate position estimation

The main goal of this step is to refine the matching result and – even more important – deliver statistically justified evidence about potential movement within the acquisition of the pan and the ms channels. Up to now, the precision of the positions in the higher and lower resolution is unknown. Therefore, the initial detections will be refined in terms of positional accuracy. Since this refinement is accomplished in a robust least square fitting approach, accuracies of the corrected position will be derived.

Assuming that active traffic moves along the road, a profile centered at the initial detection is spanned along the road direction (taken from the corresponding GIS axis). Values in the pan and similarity image are determined by bilinear interpolation for each profile point. Then, the parameters of the following Gaussian function are determined using Least Squares fitting:

$$y = s + \frac{a}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- with y ... ordinate value
- x ... abscisse value
- s ... shift in ordinate
- a ... amplitude
- σ ... width
- μ ... shift in abscisse

The accuracies of the unknown parameters (s, a, σ, μ) are obtained from their covariance matrix after an iterative fitting process. Figure 6 shows two examples of the fitted Gaussian function (continuous line) compared to the extracted profile points (dashed line). The finally refined positions in the image and the underlying profiles can be seen in Fig. 7.

2.4 Movement estimation

To make a decision about a significant movement of a vehicle, the distance between the position in the pan and the ms image is calculated. Using the known positional errors the accuracy of this distance can also be determined by simple error propagation:

$$\sigma_{dist}^2 = \sigma_{\mu_{ms}}^2 + \sigma_{\mu_{pan}}^2$$

- with σ_{dist}^2 ... variance of distance between pan and ms
- $\sigma_{\mu_{pan}}^2$... variance of position in pan channel
- $\sigma_{\mu_{ms}}^2$... variance of position in ms channels

Finally, a statistical test (i.e. Student's test) is conducted with a standard value of 5% for the probability of error.

In our tests we obtained values of less than half a pixel for the tested distances to verify a movement, i.e., 1.2m when taking the coarse resolution of the ms channels as reference scale. This corresponds to a minimum velocity of 20km/h, which is well below the typical velocity in city areas.

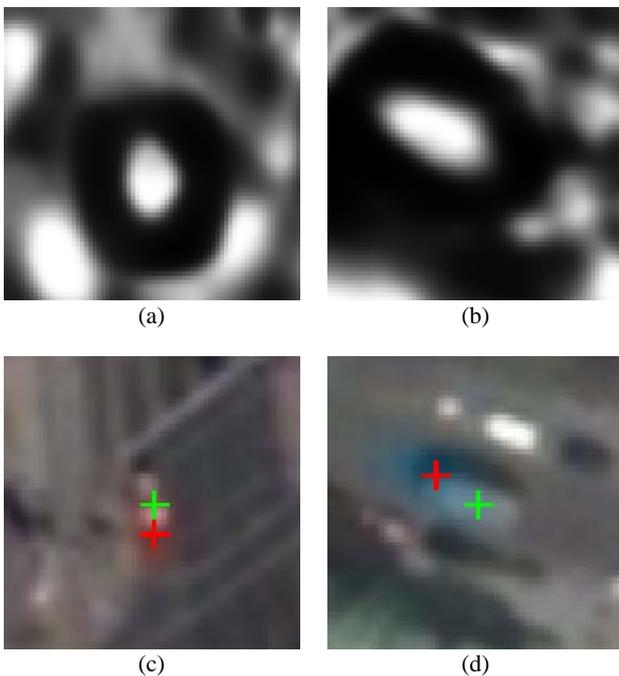


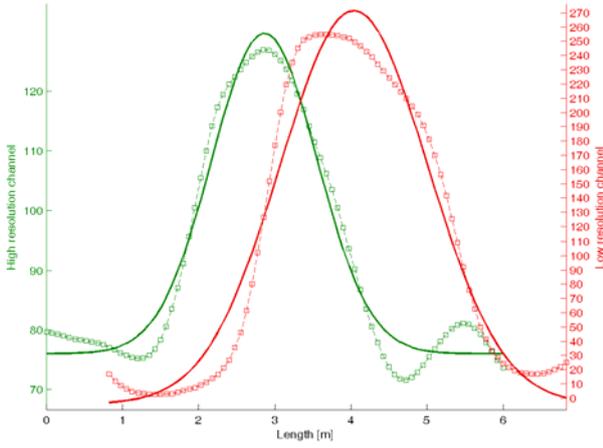
Figure 5. Matching results: (a) and (b) show similarity measure, (c) and (d) corresponding detection with given position in pan image (green) and detection in ms images (red)

3. RESULTS

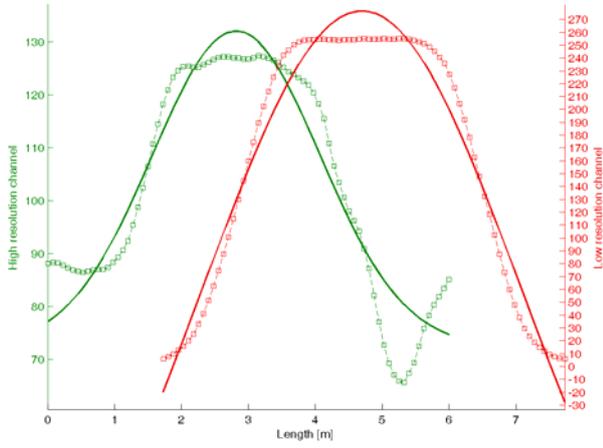
In this section more examples for the application of our approach are shown to illustrate the general performance of this approach. A numerical evaluation will be given in a later paper, once this algorithm has been added to the overall system sketched in Fig. 2.

The figures of this section are organized as follows:

- (a) original image with initial detection
- (b) similarity image from matching
- (c) results of the fitting process
- (d) final results and used profiles



(a)



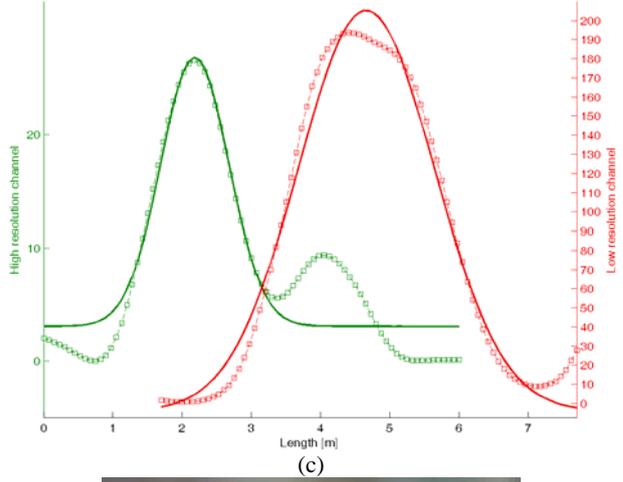
(b)

Figure 6. Refined position estimation for pan (green) and similarity (red) image. (a) and (b) correspond to the examples of initial matching in Fig. 5 (c) and (d), respectively.

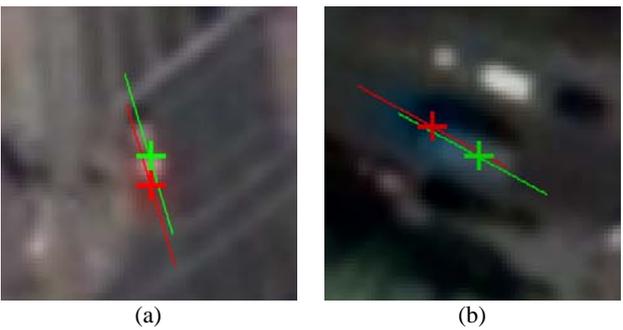


(a)

(b)



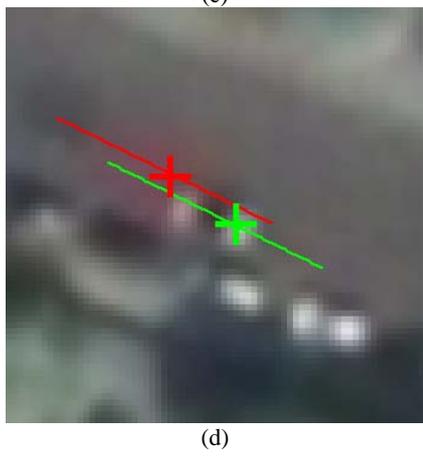
(c)



(a)

(b)

Figure 7. Profiles and accurate position for panchromatic (green) and similarity image (red). (a) and (b) correspond to the examples shown in Fig. 6 (a) and (b), respectively.



(d)

Figure 8. Example I

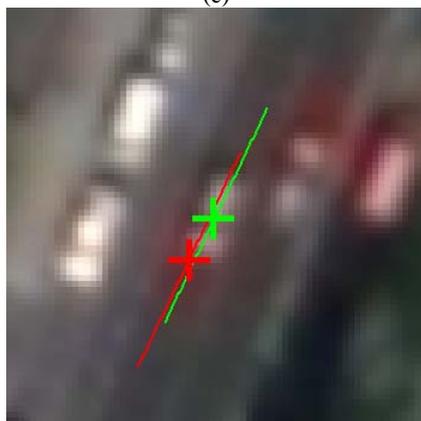
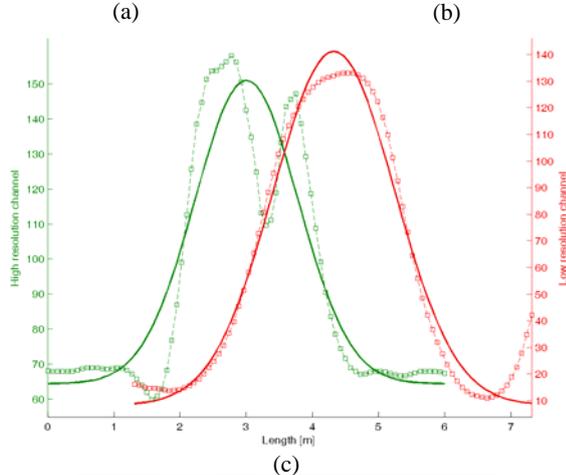
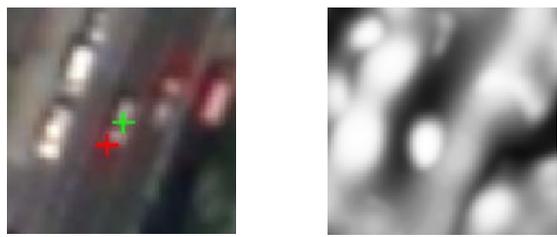


Figure 9. Example II

4. OUTLOOK

The examples show the potential of the presented approach. The position of all vehicles could be detected in the low resolution ms channels by the use of robust least square fitting in combination with a shape-based matching algorithm. This module will now be added to the overall detection for numerical evaluation. Furthermore, tests on gray vehicles which do not have large contrast in the ms channels will be performed. We expect that even in these cases movement detection can be performed, although it might be slightly less reliable.

REFERENCES

Etaya, M., Sakata, T., Shimoda, H., Matsumae, Y., 2004. An experiment on detecting moving objects using a single scene of QuickBird data. *Journal of The Remote Sensing of Japan* 24 (4): 357–366.

Etaya, M., Nakano, R., Shimoda, H., Sakata, T., 2005. Detection of Ocean Wave Movements after the Northern Sumatra Earthquake using SPOT Images, *Proceedings of IGARSS (2005)*, pp. 1420-1423.

EURIMAGE, 2007. www.eurimage.com, accessed Aug. 1st 2007.

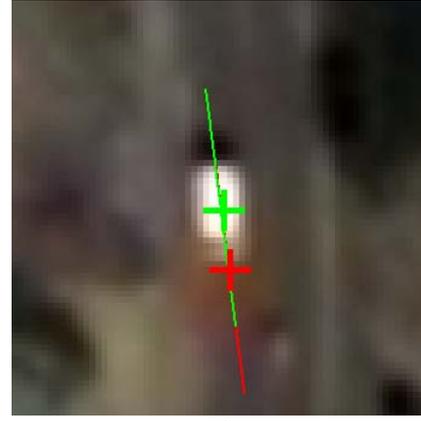
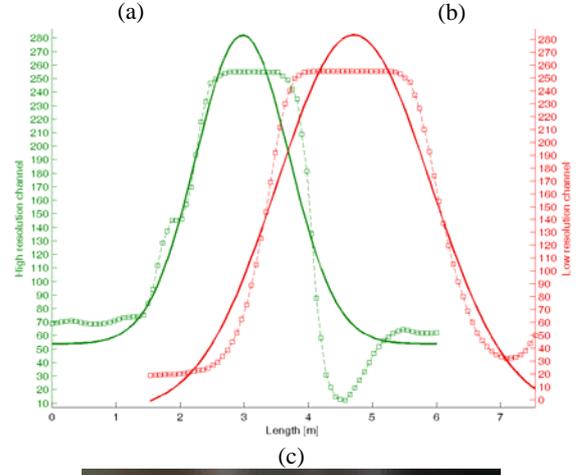
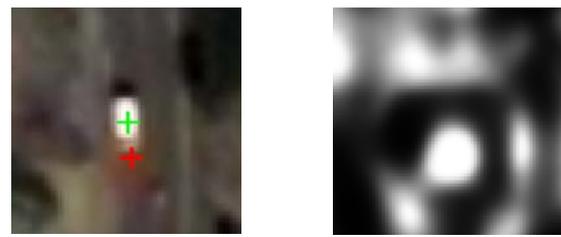


Figure 10. Example III

Hinz, S., 2005. Fast and Subpixel Precise Blob Detection and Attribution. In *Proceedings of ICIP 05, on CD*.

Hinz, S., Bamler, R., Stilla, U., 2006 (Eds.): Theme issue “Airborne and spaceborne traffic monitoring”. – *ISPRS Journal of Photogrammetry and Remote Sensing* 61 (3/4)

Hinz, S., Lenhart, D., Leitloff, J., 2007: Detection and Tracking of Vehicles in Low Framerate Image Sequences. Workshop on High Resolution Earth Imaging for Geospatial Information, Hanover, Germany

Leitloff, J., Hinz, S., Stilla, U., 2006. Detection of Vehicle Queues in Quickbird Images of City Areas. *Photogrammetrie – Fernerkundung – Geoinformation*, 4 (6): 315 - 325.

Steger, C. (2001): Similarity measures for occlusion, clutter, and illumination invariant object recognition. In: B. Radig and S. Florczyk (eds.) *Pattern Recognition*, LNCS 2191, Springer Verlag, 148–154.

Stilla, U., Rottensteiner F., Hinz, S. (Eds.), 2005. Object Extraction for 3D City Models, Road Databases, and Traffic Monitoring - Concepts, Algorithms, and Evaluation (CMRT05), *Int. Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36 (Part 3/W24)

BUILDINGS' MODELLING FOR DIGITAL ORTHORECTIFICATION OF URBAN AREAS IMAGES

B. D. Marinov

UACEG, Faculty of Geodesy, bul. Hristo Smirnenski 1, 1046 Sofia,, Bulgaria - marinovb_fgs@uacg.bg

Commission III, WG III/4

KEY WORDS: Building Modelling, Image Understanding, Digital Orthorectification, Urban Environment

ABSTRACT:

Orthophoto production is one of the main photogrammetric tasks. The usage of digital photogrammetric stations allows precise orthorectification based on the pixel resolution basis. Different data sources are used for production of DEM based on survey data, contour line maps, stereo images, RADAR or LIDAR data and combinations of them. This technique requires generation of DEM with high accuracy. One of the latest technologies for DEM creation is based on the correlation procedures of digital stereo pairs.

A two-steps procedure for processing of generated model is suggested to solve this problem. Initially the DEM is produced without taking into account the buildings. A complex model of buildings and surrounding terrain is suggested that is suitable for mapping, orthorectification and 3-D modelling. The suggested model has hierarchical structure and describes main part of buildings as walls, roofs and subparts like balconies, staircases. It is suitable for description of spatial information, oriented for digital processing of stereo images. The description of information is formulated, based on the methods of formal grammars. The several level of production rules are formulated, which are corresponding to separation of object into subparts, the generation of subparts and the producing of surfaces from graphic primitives. The software arrangement is discussed, that corresponds to application of grammar rules, which are convenient for automatic processing of digital stereo pairs. The introduced model is used for image interpretation and restoration of objects from their stereo images.

The generated model of objects in urban areas is applied for image orthorectification. The areas of buildings' roofs, walls, shadows and visible terrain were separated. The used vector model of buildings is projected over the initial images to separate them in areas of terrain, buildings, walls, invisible faces. The procedure is extended for roofs or balconies at different heights, which are overlapping in images. This allows the segmentation of DEM into areas and independent control of parameters for plane surfaces and curved surfaces or natural objects like terrain. The model of visible terrain is extrapolated over the shadow areas and the areas of buildings. After that separate models are superimposed to produce final surface model that is used for orthorectification. The modification of index matrix method is applied for orthophoto generation. The several levels mask is created. It takes into account the overlapping of building over terrain, over adjacent building and overlapping of parts of the same building. The orthophoto image is separated into parts – for buildings' roofs and walls and for surrounding terrain, and for overlapping parts. The superimposition and mosaic creation from rectified images is made as a final step of the procedure.

1. INTRODUCTION

1.1 Orthorectification of Scenes Containing Buildings

The photogrammetric generation of models from scenes containing man-made objects has its peculiarities which are similar for tasks in for orthorectification of terrain and buildings and generation of photo-realistic models in architectural photogrammetry. The common problem is the method of generation of model, its complexity and level of particularity. Main problems that have to be solved for creation of photogrammetric models of buildings are used technology schemes for generation of such models, methods for their description and generation and procedures for transformation of data format of model into format suitable for corresponding hard copy products or dynamic visualization. This requires simultaneous solving of the following three basic tasks: adequate model as combination of vector and raster data, reliability of the generated model from different sources and preliminary information for objects data sets, correspondence between model formats and visualization procedures.

Digital presentation and archiving of buildings requires simultaneous usage and data fusion from aerial or close range photogrammetry, laser scanning, digital photogrammetry and image processing techniques. Main tasks to be solved are photometric and geometric reliability between the object and

the model. The process of generation of 3D models of architectural objects requires solving the task of generation of 3D skeleton models, decomposition the images of the object into separate parts and appropriate processing of decomposed parts and adequate method for interpolation of object surfaces.

For processing of photogrammetric information is typical usage of space information. The analysis of images in urban areas is very difficult due to the buildings and artificial objects. The large displacements, hiding of terrain and hiding the parts of adjacent objects and influence of the shadows generate very complicated images. Similar problems arise for processing of images in close range photogrammetry for which there are important protruding parts of objects. This produces problems for automatic identification of corresponding parts in stereo images of such objects or territories.

The influence of quality of DEM over the geometric precision of orthoimage is analysed in (Pala V., Albiol R., 2002). A geometric correction algorithm is applied for combination of separate parts of orthoimages from different views of the same area to solve the problems with the hidden parts of terrain. Importance of including DEM in orthorectification process is analysed in (Jauregui M., Vilchez J., Chacon L., 2000).

Usage of digital building model for accurate orthophoto generation is presented in (Bock-Mo Y., Eui-Myoung K., In-

Tae Y., 2000). The process of orthorectification is preceding by automatic building extraction.

In semi-automatic procedures for orthorectification of building images it is possible excluding of special kind of lines for such objects – like border lines, which allow excluding of unwanted parts of object model. The structure of spatial information is very important for reliability and speed of processing in such systems.

1.2 Automatic Building Reconstruction

Building detection and reconstruction is based on different techniques and uses wide variety of procedures. Different classification schemes are suggested for image processing and interpretation, taking into accounts the shape and height characteristics of buildings (Weidner U., Förstner W., 1995). The semi-automatic procedures for building detection are developed, which are based on LSQM matching, usages the geometric characteristics in object space (Zhang, Z., Zhang, J., 2000). Involving the geometrical constrains of building roofs gives possibilities for semi-automatic building extraction. There are a lot of investigations in which the shadow information is connected with geometric characteristic of objects.

The building analyses process is based on object models, which are using the verification of feature sets (Ameri, B., 2000). The development of this idea is realized in procedure taking into account the topological and geometric characteristics of objects in reconstruction of regular shape faces, which approach is implemented in (Heuel, S., Förstner, W., Lang, F., 2000). Some problems of structural approach to man-made objects are discussed in (Michaelson, E., Tilla, U., 2000). Model based methods are considered using geometric, topologic and structural knowledge based on polyhedral parametric or generic models. Increasing the accuracy of processing is achieved by applying of data base for roof corners (Chio, S.H., Wang, S.C., Wrobel, B., 2000). The procedure for building extraction from high-resolution Digital Surface models containing terrain and buildings information is suggested in (Weidner U., 1996). The building reconstruction is based on the usage of parametric and prismatic building models. There are developed integrated algorithms, which apply data fusion of stereo images, GIS data or digital map information (Suveg, I., Vosselman, G., 2000).

The down to up procedure for building detection is introduced in (Zhao, B., Trinder, J.C., 2000). The hierarchical approach consists of three major stages: building detection, building segment extraction, 3D feature matching and building modelling. The hierarchical approach for building reconstruction using 3D information, such as color, texture, or shadow and reflectance is introduced by (Seresht M.S., Azizi A., 2000). The structural approach for generation of object model from aerial images of urban scenes is suggested in (Stilla U., Jurkiewicz K., 1996). A bottom to up solving procedure is tested for these purposes.

Utilization of reliable procedure for building extraction and reconstruction which is compatible with processed images is very important for quality of orthorectification process.

2. MODEL STRUCTURE

2.1 Image Description

Main method of description is based on the topological properties of areas in the stereo images. A picture language grammar technique could be applied to formalise the process of image description and to cover different variants of image

configurations. Such approach is very large used for description of two dimensional or three dimensional images [Fu, 1982]. The formalised the process of description of contour images and generation of hierarchical description of image has to be enlarged for images of 3D objects and to involve additional information. The information that has to be taken into account includes the parallax of contour lines, the position of object relatively to centre of the photo and topological properties of areas in separate images. In this situation only the walls oriented to the centre of photo could generate visible areas in image. The main properties of image description have to take into account the relation between areas into image and connection between areas and contours. The arcs are used as main primary elements for description. Arcs present the segments of lines between two points of connection. From topological point of view such points are only points of connection of more than two arcs by taking into account the type of arcs (wall edges, foundation border and so on). By this reason the segmentation of arcs has to be applied. The main topological properties that have to be taken into account are the relation of adjacent areas, clusters of areas, included clusters of areas, hanging areas (with only one point of connection). To take into account the specific properties of artificial objects have to be involved areas of types roofs, walls, shadows, hidden terrain or roof areas).

For purposes of contour description only the tie points and arcs could be used. In such situation it would be difficult to generate relation between areas. Usage of only topological properties of areas and their relation would increase difficulties in contour generation and arcs processing.

In general case the solution is very complicated but applying the restriction of relatively small angles between the projection rays and vertical planes (as walls) the presentation could be simplified. Taking into account such restriction the following groups of arcs could be formulated:

1. Arcs without changes in height between two areas or with relatively low difference in slope;
2. Arcs with jump in height between two adjacent areas;
3. Arcs defining break in surface slope, not in height.

In situation if nodes are marked only at that points where more than two arcs are connected some combination of height arcs could appear and should be added to terminal arcs set.

It is necessary to be mentioned that not all topological types for different height types of arc exist. According to this the following extended set of topological and height types of arcs are possible that are forming the set of finite elements of grammar.

1. Non connected elements;
2. Second order connected elements (one pointer to inside cluster list and one implied connection to own contour);
3. Third order connected elements containing two pointers to the next arcs of the own and adjacent contour and one pointer to main arc of the own contour, or to the main arc the upper level contour for element of type d;
4. Fourth order connected elements (two pointers to internal and external contour and one pointer to internal isolated cluster list and one implied connection to own contour);
5. Fifth order connected elements (two pointers to internal and external contour, one implied connection to own contour, one pointer to internal isolated cluster list and one pointer to the next cluster).

2.2 Description in 3D Space

The basis of description is presentation in 3D space, i.e. the presentation of objects as 3-dimensional. The topological and geometrical dispositions are substantial. For topological description it is applied a hierarchical structure, which is based on dividing the objects into object groups. Every object group can contain the adjacent set of objects (cluster of objects) or area of sub-objects, which distributed regularly or randomly. At the lowest level are used sub-objects, which are described by their surfaces. The object groups and sub-objects contain information about the relation and connection between them. Two main groups of factors are taken into account – the possibility for storage the description into data base and formal structural description and its generation by picture grammar. The corresponding tables forming structural description are separated in following groups: Objects, Sub-objects, Faces, 3D Arcs and Nodes. Such description is presented in (Marinov B.D., 2002). Extension of this idea to cover description of group of buildings and the objects that consists of sets of objects groups. A simple description of building object includes: roofs, walls, foundation and stairs.

Object coordinates define the base point position in coordinate system of scene (area of description). Object consists of parts named ObjGroup. Object Group may contain another Object groups or SubObject. The quadruple tree is used for production of this description.

The table of sub-objects contains information about the object and its neighbours. They are described by their faces. The table has the form

Arc table contains information for arcs with two faces from both sides of the arc:

Table defining nodes has the following form:

There are additional tables, which are describing coordinate information. They are vectors of origin (Position) and orientation (Rotation). Another two tables describe surfaces and arc segments. (Surfaces, Segments).

Such organization of information is compromise between memory requirements and speed of data access. The cluster arrangement of information requires more memory allocation for description but increase the speed of processing. This organization is adopted for presentation analyses and presentation of space information for object and sets of objects. Larger segment size increases the speed of processing but consumes more memory for short segments.

3. FORMULATION OF PICTURE GRAMMAR FOR MODEL DESCRIPTION

3.1 Picture Description Languages

Picture description languages are widely used for presentation of 2D and 3D objects and their images (Fu, K.S., 1982). The necessity of two dimensional element descriptions involves usage of extended type of grammar. Suitable for that purposes is PLEX-GRAMMAR [Feder, 1971]. A N-attaching-point element (NAPE) is introduced in it. The grammar is represented by the six-tuple $(V_T, V_N, P, S_0, Q, q_0)$, where

V_T - is a finite non-empty set of terminal (non-productive) elements;

V_N - is a finite non-empty set of non-terminal elements (productive);

$$V_T \cap V_N = 0; \quad (1)$$

P - is a finite set of productions (generating rules);

S \in V_N is a special NAPE - initial;

Q is a finite set of symbols called identifiers that form joint lists (of internal joint points of NAPE set) and tie-point list (consisting of external points of NAPE set) - identifying the links between elements. The identifiers could be represented by integer corresponding to the number of tie-points of single NAPE.

$$Q \cap (V_T \cap V_N) = 0; \quad (2)$$

$q_0 \in Q$ is a special NULL identifier that is used to show that corresponding NAPE from NAPE list is not connected to the described joint or tie point.

An extension of grammar is used in which except traditional joint point for which are necessary at least two non-zero element in joint list a special type of joint is used for contour definition joint which could involve only one connected non-zero element, belonging to the main arc of the contour. According to the above definition a set of four NAPEs with five joint points (four for arc connection and one for contour definition) and two tie-points is represented by set of components and two lists (of joint and tie-points) of the form:

$$V_1 V_2 V_3 V_4 (q_1 q_2 q_3 q_4 \cdot q_1 q_2 q_3 q_4 \cdot q_1 q_2 q_3 q_4 \cdot q_1 q_2 q_3 q_4; \quad (3) \\ q_1 q_2 q_3 q_4) (q_1 q_2 q_3 q_4 \cdot q_1 q_2 q_3 q_4)$$

3.2 Picture Grammar for Projective Images of Buildings

Defining of picture grammar with such type characteristics, which satisfies generation of projective images of buildings is formulated in (Marinov B.D., 1996) and is further developed in (Marinov B.D., 2004). The following set of arcs is defined.

1. Arc without height change on two sides of areas.

m – border of different cover;

s – shadow over terrain, wall or roof.

2. Arc, which separates two areas with different heights (roof-terrain, roof-wall, roof-roof).

r – border roof/terrain (roof/roof);

v – border between wall edge/terrain (wall edge/roof);

p – border roof / wall;

w – border wall / wall (between different buildings or their parts).

3. Arc, which divides two faces with different slope of surfaces.

h – edge wall / roof (at eaves);

t – edge terrain / wall (roof / wall in cases when part of building lies over the lower roof or terrace);

e – visible edge between surrounding walls;

k – edge between different roof faces.

It is possible to add combined types, which include sequence of two arcs with different heights.

4. Arcs, which are combination of two arc segments: j, l (wall arcs), f, g (roof arcs).

5. Arcs, which are combination of three segments: u (wall), n (roof).

Definition of different type of generation rules requires additional indexes for main or auxiliary arcs.

K, M, C, D – main external arcs of last or non last contour;

B, S – auxiliary arcs of external contour or different level contours;

I – internal arc of contour at same level;

O, Q – single or isolated contours.

Identification of different state of generation is made by second index that shows the specific state of the arc:

ρ - state of generation of isolated contour;

Ω - state of generation of external contour arc;

χ - generation of hidden contours;

and states for arc shift - $\Sigma, \Psi \Phi$.

3.3 Producing Rules Examples

The following examples for producing rules include several main groups of rules. The first group defines generation of contours and set of isolated contours.

$$I_0() \rightarrow i_E() \quad I_Q() \rightarrow I_Q A_O(11)(02) \\ A_O(1,2) \rightarrow A_{O\rho}(1,2) \quad R_O(1,2) \rightarrow R_{O\rho}(1,2) \quad (4)$$

Generation of internal contours begins from main contour or from main arc in state ρ :

$$Y_{O\rho}(1,2) \rightarrow Y_{O A_O}(21)(10,02) \\ W_{K\rho}(1,2,3,4) \rightarrow W_{K A_O}(41;30)(10,20,02) \quad (5)$$

where symbol $Y \in (A, R)$ и W is arbitrary arc from set of internal contours (A, R, U, J, V, T).

The second group of rules is connected with generation of adjacent contours by applying of rules for generation of surrounding arcs. Initial set of producing rules includes generation of adjacent contours of type A or Y.

$$A_0(1,2) \rightarrow Y_I A_K Y_C(011,102,220;330,003) (040,004) \\ A_Q(1,2,3) \rightarrow Y_I A_M Y_C(011,102,220;330,003) (040,004,050) \quad (6)$$

The separate set of rules produces external or internal self-closed contours. External adjacent contours are produced by following rules:

$$A_O(1,2) \rightarrow A_K Y_C(12,21;30,03)(40,04) \\ A_Q(1,2) \rightarrow A_M Y_C(12,21;30,03)(40,04,050) \quad (7)$$

The rules for generation of internal isolated contours have similar form.

$$A_O(1,2) \rightarrow A_D Y_K(11,22;30,03)(40,04) \\ R_O(1,2) \rightarrow R_D Y_C(11,22;30,03)(40,04) \quad (8)$$

The generation of moving arc begins from node between arcs or from internal arc point. For arcs of type A they have the form:

$$A_X \rightarrow A_X A_I \Sigma Y_C \Psi A_X \quad W_X A_X \rightarrow W_X Y_I \Sigma \\ A_X \rightarrow R_X R_I Y_C \Omega \quad W_X A_X \rightarrow W_X Y_I \quad (9)$$

Rules for arcs of type R are identical.

Additional set of arcs define shift of end of external arc. Another set of rules are connected with generation of external visible walls and finalizing the producing sequence in cases when they are independent or their images lie over the parts of adjacent objects. The rules for transformation of including or surrounding contours are added. Another set of rules transforms external into internal contours. Two stages transformation procedure are used, which modifies arc types.

Such construction of grammar allows modelling the hiding of building parts depending of view point and the shadow influence and their covering. This procedure is applied at topological level.

4. RELATIONS BETWEEN MODEL PARTS IN 3D BUILDING MODELS

In three-dimensional model there are only relations of connectivity between areas. The corresponding 2-D model has adjacent and overlapping areas. For modelling of overlapping areas it is necessary to extend the possible relations between types of areas with different lines. The application of formal picture languages is appropriate techniques for analyses of images of such object configurations (Marinov, B., 1996).

The simple model of terrain and buildings consists of terrain and buildings. The buildings are constructed from walls and roof. More sophisticated model includes balconies on the walls, terraces on the top roof and tower over the roof and adjacent

buildings or parts of them with different height. To define this model the following groups of plane objects can be formulated:

A. Top covers: roof, peak, floor, terrain (ground);
B. Side faces: wall, parapet, cornice, internal wall, curved shells

C. Bottom covers: foundation, eaves, ceil.

D Internal faces (invisible)

In space model of objects there are different combinations of surfaces to produce whole object.

For nodes of order four the combination are more complicated but from practical point such nodes are not so much. Some examples are: N (roof, roof, cornice, cornice); N (roof, roof, wall, wall); N (roof, roof, roof, roof); N (wall, wall, sill, cornice).

Main production rules generate building's body, roof, surrounding walls. Other types of rules include production of objects that are attached to the surfaces of main object or sub-object. Such objects are chimneys, mansards, stairs, platforms, shelters, and bays. The special types of objects are groups of columns or monuments under external shelters or internal bays. Some examples of different types of faces are shown in figure 1

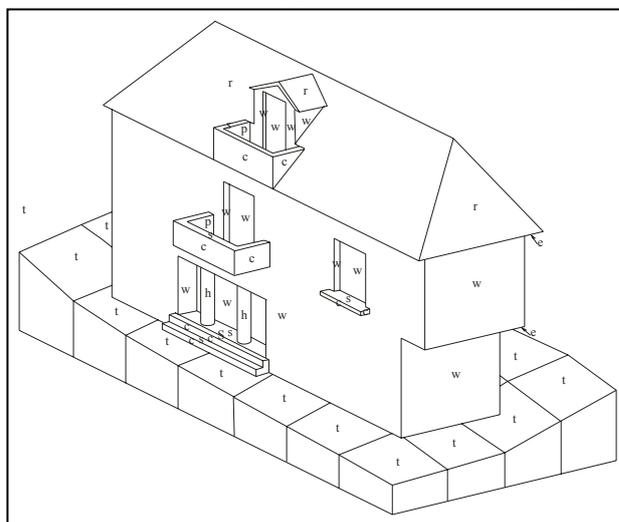


Figure 1. Examples of some face types

It is important to note that generation of nodes of specific type is feature of defined production rule and that usually the production rules of specific types generate usually couples of nodes. It is important for reconstruction of object model in cases when not all nodes are visible. This requires modules for analyses of corresponding types of nodes and production of invisible nodes. Such processes arise in mapping of architectural objects when some nodes are not visible.

Edges between not adjacent faces create another type of problems. The main types of combinations include roofs, walls, shadows and terrain (Marinov, B., 2002). More extended set of combinations include different protruding objects or sub-objects.

Taking into account of these relations is very important for analyses of data and for control of data coding. From another side including of this information in data stream will create additional efforts for automatic recognition of types of boundary and types of nodes. This requires definition of data structure for input of photogrammetric information that is suitable for automatic of operator controlled data input.

Based on this model the following extensions for description of 3D information can be suggested:

1. The hieratical structure must include the conjunction faces between main object and sub-object.

2. For every planar face the orientation vector of surface must be calculated and added as attribute information.
3. For every object or sub-object is defined and calculated the surrounding parallelepiped body, which includes entirely object.
4. For every arc are added as attributes the identifier codes for two adjacent surfaces.

This structure is convenient for solving the tasks for visibility and for creation of index matrixes for orthophoto transformation in photogrammetry.

5. PRESENTATION AND FORMING OF ORTHOPHOTO BASED ON BUILDING MODEL

The advantage of suggested complex model is the possibility to decrease the number of checks or visibility of processed pixel. To solve the problem a surrounding parallelepiped bodies are used.

The results for one complex building (church) are shown on the following figures. The digital images with superimposed vector model are shown in figure 2.

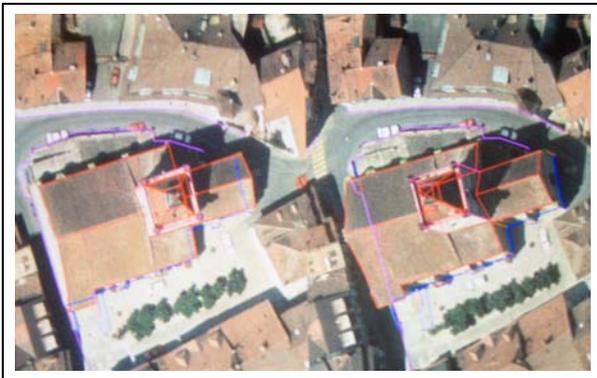


Figure 2. Source stereo couple with vector model

The basic part of the orthoimage is produced from left image.



Figure 3 Orthorectified image generated from left image

The problem of generation of full model is that some lines are not visible in stereo or are not visible. For such lines the fictitious addition of model has to be generated.

The reverse scheme of orthorectification is used. The check for visibility of pixels is made based on calculation the limits of parallelepiped bodies surrounding every sub-object. For

projection ray from pixel in orthoimage corrected by height from model is checked the visibility by comparison with limits of surrounding bodies of adjacent parts of same building or nearest buildings. If tested pixel is not visible in current digital image the test for coupling image is made. It is possible to make checks not only in strip direction – longitudinal overlap but also for side overlap of images.

Additional part of orthoimage is produced from another stereo image. The processing of pixels from other images is made only for hidden pixels in base image. It is possible to generate separate orthoimages from every of neighbouring image. Which pixel is used is coded in extended index matrix. The used codes are visible terrain, visible roof, visible balcony for every different visible surface. The index matrix for usage of pixels is generated during this process. The extension of index codes includes not only invisible pixels of terrain but also identification number of slope roofs, flat roofs or external stairs and balconies. The generated codes allow usage the most appropriate nearest pixels of visible parts: terrain, roofs, balconies and corresponding codes for surfaces of invisible pixels. Additional part of index code is number of image from which the source data are taken.

The hidden parts from right image are shown in figure 4.

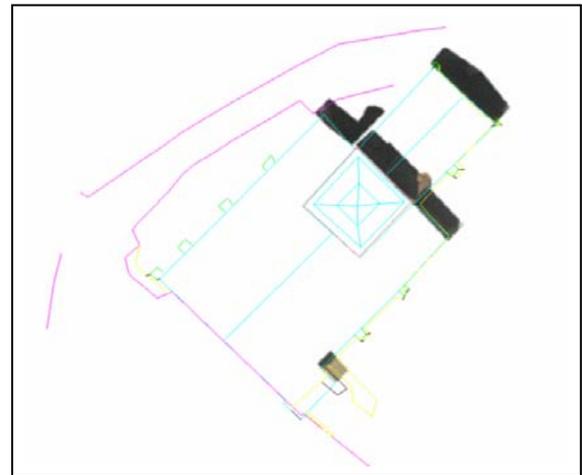


Figure 4. Ortho chips from right image

The final orthoimage produced as combination of two orthorectified images with superimposed vector model is shown in figure 5.



Figure 5. Composition of orthoimages from suitable parts of adjacent images

Substantial advantage of this model is possibility for simultaneous analyses of two images of stereo couple and search of corresponding areas based on structural description. This allows enhancing the reliability and accuracy of obtained information.

6. DISCUSSIONS AND CONCLUSIONS

The suggested method for description of 3-D spatial data information is suitable for solving the task for visibility analyses in GIS and for orthophoto production in Digital Photogrammetry. The main problems are level of complexity of the model, the structural extension of model for invisible surfaces and lines for which the virtual lines are produced to ensure the concordance of the model. The computing efficiency for generation of orthoimage depends on additional information for consisting parts of object and the nearest object which can influence the pixels in the neighbouring objects.

It is possible to be used in cases of different resolution of model with variable degree of generalization. This model is suitable for objects with different complexity and particulars.

The suggested method enlarges the possibilities for DEM generation and orthorectification with systems for digital photogrammetry for urban areas. The suggested method could be used in procedures for archiving and dynamic visualization of architectural objects. The produced orthophoto is more accurate and adequate of the objects that are presented on it. As a final result the technology based on orthophoto production and digitising of accurate orthophotos could be applied successfully for urban areas mapping and planning.

REFERENCES

- Ameri, B., 2000. Feature Based Model Verification (FBMV): A New Concept for Hypothesis Validation in Building Reconstruction. In: *IAPRS*, vol. XXXIII, Part B3, Amsterdam, pp. 24-35.
- Bellman C.J., Shortis M.R., 2004. A Classification Approach to Finding Buildings in Large Scale Aerial Photographs. In: *IAPRS (XXth Congress of ISPRS)*, vol. XXXV, Part B3, Istanbul, Turkey, pp. 337-342.
- Bock-Mo Y., Eui-Myoung K., In-Tae Y., 2000. Generation of Digital Orthophoto for Urban Area Using Digital Building Model. In: *IAPRS, (XIXth Congress of ISPRS)*, vol. XXXIII, Part B4, July 2000, Amsterdam, Netherlands, pp. 462-468.
- Chio, S.H., Wang, S.C., Wrobel, B., 2000. Interactive Roof Patch Reconstruction Based on 3-D Linear Segments. In: *IAPRS, (XIXth Congress of ISPRS)*, vol. XXXIII, part B3, Amsterdam, Netherlands, pp. 183-190.
- Heuel, S., Förstner, W., Lang, F., 2000. Topological and Geometrical Reasoning in 3D Grouping for Reconstructing Polyhedral Surfaces. In: *IAPRS, (XIXth Congress of ISPRS)*, vol. XXXIII, Part B3, Amsterdam 2000, pp. 397-404.
- Fu K.S., 1982. *Syntactic Pattern Recognition and Applications*. Prentice-Hall, Englewood, Cliffs, New Jersey, pp.129-162.
- Feder, J., 1971. Plex Languages. *Information Science*, 3, pp.225-241.
- Jauregui M., Vilchez J., Chacon L., 2000. Digital Orthophoto Generation. In: *IAPRS, (XIXth Congress of ISPRS)*, vol. XXXIII, Part B4, Amsterdam, Netherlands, pp. 400-407.
- Marinov B., 1996. Processing of Protruding Objects in Digital Images of Urban Areas. In: *IAPRS*, vol. XXXI, Part B3, Vienna, pp. 497-502.
- Marinov B., 2002. Structure of 3-D Information for Processing of Digital Stereo Images. In: *Proc. of the Jubilee Scientific Conference "60 Years UACEG"*, vol. 8, Sofia, November, pp. 6978. (in Bulgarian)
- Marinov B.D., 2004. Structure and Usage of Photogrammetric Information in Spatial Information Systems. In: *Papers on International Symposium: "Modern technologies, education and professional practice in geodesy and related fields"*, 4-5 November, Sofia, Bulgaria, pp. 213-222.
- Michaelsen E., Stilla U., 2000. Problems in Geometric Modelling and Perceptual Grouping of Man-Made Objects in Aerial Images. In: *IAPRS, (XIXth Congress of ISPRS)*, vol. XXXIII, part B3, Amsterdam, Netherlands, pp. 577-583.
- Pala V., Albiol R., 2002. True Orthoimage Generation in Urban Areas. In: *Proc. of the Third International Symposium Remote Sensing of Urban Areas*, vol. 1, Istanbul, Turkey, pp. 309-314.
- Seresht M.S., Azizi A., 2000. Automatic Building Reconstruction from Digital Terrain Images. In: *IAPRS, (XIXth Congress of ISPRS)*, vol. XXXIII, Part B3, Amsterdam, Netherlands, pp. 792-798.
- Stilla U., Jurckiewicz K., 1996. Structural 3D-Analyses of Urban Scenes from Aerial Images. In: *IAPRS (XVIIIth Congress of ISPRS)*, vol. XXXI, Part B3, Vienna, Austria, pp. 832-838.
- Suveg I., Vosselman, G., 2000. 3D Reconstruction of Building Models. In: *IAPRS, (XIXth Congress of ISPRS)*, vol. XXXIII, Part B2, Amsterdam, pp. 538-545.
- Weidner U., Förstner W., 1995. Towards Automatic Building Extraction from High-Resolution Digital Elevation Models. In: *ISPRS J. of Photogrammetry and Remote Sensing*, 50 (4), pp.38-49.
- Weidner U., An Approach to Building Extraction from Digital Surface Models, In: *IAPRS (XVIIIth Congress of ISPRS)*, vol. XXXI, Part B3, Vienna, Austria, 1996, pp. 924-929.
- Woestyne I.V.D., Jordan M., Moons T., Cord M., 2004. A Software for Efficient DEM Segmentation and DTM Estimation in Complex Urban Areas, In: *IAPRS (XXth Congress of ISPRS)*, vol. XXXV, Part B3, Istanbul, Turkey, pp. 134-139.
- Zhang Z., Zhang J., 2000. Semi-automatic Building Extraction Based on Least Squares Matching with Geometrical Constraints in Object Space. In: *IAPRS, (XIXth Congress of ISPRS)*, vol. XXXIII, Part B3, Amsterdam, pp. 1022-1025.
- Zhao, B., Trinder, J.C., 2000. Integrated Approach Based Automatic Building Extraction. In: *IAPRS, (XIXth Congress of ISPRS)*, vol. XXXIII, Part B3, Amsterdam, pp. 1026-1031.

IDENTIFICATION OF URBAN FEATURES USING OBJECT-ORIENTED IMAGE ANALYSIS

O. D. Mavrantza^a and D. P. Argialas^b

^aKTIMATOLOGIO S.A. (Hellenic Cadastral Agency), Messogion Ave 339, Halandri, 15231, Athens, Greece – omavratz@ktimatologio.gr

^bLaboratory of Remote Sensing, School of Rural and Surveying Engineering, National Technical University of Athens, Greece, Heron Polytechniou 9, Zografos Campus, 157 80, Athens, Greece – argialas@central.ntua.gr

KEY WORDS: Remote Sensing, Buildings, Edge Extraction, DSM, Automation, Land cover, Knowledge-based System

ABSTRACT:

The objective of this work was the presentation of a knowledge-based framework for the identification of urban features, as well as additional land cover types, contained in the image scene. The test area used was located in an urban environment of Aghios Stefanos, Attica. An IKONOS image (spatial resolution of 1m-resampled) was acquired for the study area and further pre-processed. Furthermore, supplementary geo-data were created and embedded in the knowledge-based system focused on the identification of urban characteristics (e.g. building boundaries), such as derived thematic data from remote sensing techniques and the constructed DSM of the study area (with a spacing of 3m), which was extracted using photogrammetric methods, for providing the required 3-D information.

The knowledge-based methodological framework consisted of: (1) Problem conceptualization and knowledge acquisition, which involved the identification of urban features and surrounding land cover classes faced by “decoding” of knowledge based on the existing data types and context. Furthermore, remote sensing analysis methods, implementation of edge detection algorithms and multi-scale hierarchical segmentation were investigated and applied. (2) Knowledge representation, which was performed using object-oriented image classification using fuzzy representation models (proper Fuzzy Membership Functions (FMF) or Nearest Neighbor Classification (NNC)) for the assignment of objects into the defined classes.

The output result of the proposed knowledge-based identification system is expected to be a final map that contained urban features of interest (such as building boundaries and road segments) and additional land-cover classes which consist the natural scene of the study area (e.g. vegetated regions).

1. INTRODUCTION

Photointerpretation of linear information is a subjective process and therefore there is a substantial need for automation of extracting linear information using automated techniques. Certain efforts were made in this direction including the application of edge enhancement and parametric / non-parametric detection algorithms, segmentation algorithms, wavelets, variational models, active contours, neuro-fuzzy systems, Hough Transform etc.

However it is difficult to choose among optimal algorithms since the complex scenes portrayed on satellite images are strongly dependent on the radiometric and physical properties of the sensors and on the illumination properties and topographic relief of each scene. Usually, the nature of information to be extracted and its scale and context determines the “suitability” of the method applied for linear feature extraction (medium-level analysis).

Concerning high level analysis (image understanding), the identification of the extracted features consists a more complex research topic due to limitations of the designed recognition systems as well as due to the incompetence of fully “simulating” human perception in different recognition tasks. Much scientific effort has been made in the domain of image understanding, where several recognition methodologies have been proposed for serving diverse application tasks (e.g. biomedicine, understanding of natural scenes, etc). Integrated

image analysis systems involving multi-scale and multi-source data segmentation, object-oriented representation, fuzzy and other types of classification have been recently involved for various applications. A similar methodology by the authors has successfully been applied for the identification of topographic and geological lineaments using coarser spatial resolution satellite data (such as LANDSAT-ETM+, with a spatial resolution of 30m) in a sedimentary, not urban, geotectonic environment (Mavrantza and Argialas, 2003, 2006).

1.1 Edge Detection Operators: Overview

In image processing and computer vision, edge detection treats the localization of significant variations of a gray level image and the identification of the physical and geometrical properties of objects of the scene. The variations in the gray level image commonly include discontinuities (step edges), local extrema (line edges) and junctions. Most recent edge detectors are autonomous and multi-scale and include three main processing steps: smoothing, differentiation and labeling. The edge detectors vary according to these processing steps, to their goals, and to their mathematical and computational complexity (Ziou and Tabbone, 1997).

In the present work, only step edge detectors were examined, which can generally be grouped into two major categories:

1. Optimal gradient-based detectors (e.g. the Canny algorithm, etc.).

2. Operators using parametric fitting models (e.g. the *EDISON* edge detector by Meer and Georgescu (2001), etc) (Ziou and Tabbone, 1997).

1.2 Urban Feature Extraction and Mapping: Overview

Concerning the automatic approach in urban feature extraction (specifically, roads and building boundaries), several categories of processes were applied including:

1. Diverse wavelet analysis methods such as edge detection based on scale multiplication (Zang and Bao, 2002), contourlet transform (Do and Vetterli, 2005), etc.
2. Snakes (Agouris *et al.*, 2001), Variational models, geodesic active contours (Mumford and Shah, 1989, Karantzas, *et al.*, 2007).
3. Optimal edge detection algorithms (Heath *et al.*, 1997).
4. Object oriented image analysis for the identification of informal settlements (Hoffman, 2001).

1.3 Motivation and aim

From the thorough examination of the literature (as it was also presented in Section 1.2), it is inferred that computer-assisted methods for the extraction of urban features such as building boundaries and road segments were mostly based on the spectral and textural characteristics of these features.

Despite of the good spatial accuracy most of the aforementioned methods provide, redundant non-urban features with similar spectral characteristics were also extracted. In such cases, additional post-processing approaches are required for the discrimination of the urban and the non-urban features (Argialas, *et al.*, 2007).

On the other hand, optimal edge detectors (e.g. the Canny algorithm) have already been successfully applied on multi-scale natural scenes with quite satisfactory results (binary images with one-pixel thickness, efficient length, pixel connectivity and very good spatial (even sub-pixel) accuracy) (Heath *et al.*, 1997, Mavrantza and Argialas, 2007). The integration of optimal edge detection techniques and a knowledge-based approach would quite sufficiently combine the textural and spectral information of all the intrinsic features of a natural scene portrayed in a satellite image with the spatial and contextual information of their adjacent features for the discrimination of their semantic content. That means the discrimination of urban and non-urban features can be achieved by exploiting the semantic information of the participating features.

The integration of optimal edge detection techniques and a knowledge-based approach has already been investigated by the authors for the extraction and identification of topographic and geologic lineaments with quite successful results (Mavrantza and Argialas, 2006).

2. METHODOLOGY AND RESULTS

2.1 Study area and data used

In the present study, an IKONOS very high resolution (VHR) image of the extended area of Attica Prefecture (Aghios Stefanos area), with a spatial resolution of 1m-resampled and

acquired in year 2000 (Figure 1), as well as the Digital Surface Model of the same area with 3m spacing were used as initial input geo-data.

In Figure 1, the pseudocolor composite RGB-421 of the IKONOS VHR multispectral image with spatial resolution of 1m-resampled and size 934x1070 pixels is presented. Man-made features appear with hues of blue, while vegetated areas appear with different hues of red due to the vegetation reflectance in the infrared band.



Figure 1: Pseudocolor composite RGB-421 of the IKONOS VHR image with spatial resolution of 1m-resampled and size 934x1070 pixels.

2.2 Image pre-processing

In the pre-processing stage, the IKONOS of the study area was geodetically transformed into the Transverse Mercator Projection and the Hellenic Geodetic Datum (HGRS87). The positional accuracy of the georectified image was approximately from 1.5-3.0 meters.

For the implementation of the Pratt evaluation metric (which is in detail described in a following section), an ancillary ground truth (reference) file was required as input. This ground truth file was developed containing all the visually interpreted linear segments related to building boundaries as well as road segments, from the IKONOS image (and verified on the ground), represented with their X, Y coordinates and the total number of the actual edge points (in an ASCII format file).

2.3 Framework for the design of the knowledge-based system

The present work is a generalized framework of the design of a knowledge-based system for the automated extraction and identification of urban / peri-urban features and additional land cover types of the natural scene (Figure 2).

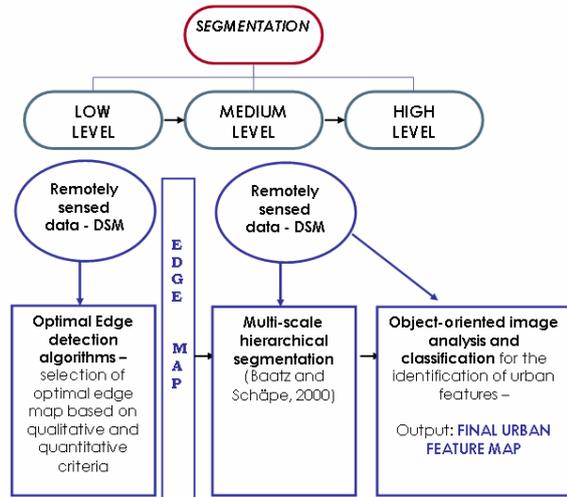


Figure 2: General methodological framework for the extraction and identification of urban features (Mavrantza and Argialas, 2007).

The knowledge-based framework consisted of the following methodological stages:

1. Problem conceptualization and knowledge acquisition: At this stage, the identification of urban features and surrounding land cover classes was faced by covering aspects like (a) the determination of the type of acquired and created geo-data in order to assist to the derivation of the information of interest, (b) the categories of expected land cover to be identified, (c) the sources of information to be required for feature identification (e.g. photointerpretative (qualitative) knowledge). In addition, “decoding” of knowledge was performed based on the existing data types and context. Furthermore, low- and medium level image analysis was performed in order to obtain the required geo-data for being introduced to the knowledge-based identification system. Low-level image analysis techniques contained (a) *Remote sensing analysis methods* and (b) *Edge detection algorithms*. The implementation of optimal edge detection algorithms (e.g. the Canny algorithm, the *EDISON* algorithm, etc) as well as their qualitative / quantitative assessment, were performed for the creation of the “optimal edge maps”, which contained the highest amount of information of interest and at the same time shall keep the false-positive (redundant edges) at the minimum level. These output edge maps were further introduced into the knowledge-based system for being “transformed” into the final thematic (urban feature) map containing the identified features of interest. Finally, medium level image analysis contained multi-scale hierarchical segmentation (Baatz and Schäpe, 1999) was also performed in order to create object segments to be assigned into semantic land-cover classes in the knowledge-based system.

2. Knowledge representation: At this stage, knowledge representation was performed using object-oriented image classification using fuzzy representation. The inherent land cover classes / subclasses were determined at each level of hierarchy according to the type and information content of geodata used. The determination of the proper class / sub-class attributes was based on the following intrinsic image properties: *spectral, geometric, spatial relation and context*. These attributes followed the “general-to-specific” inheritance principle. In addition, the class attributes were connected with *AND / OR / MEAN* logical operators according to each weight

criterion for the further assignment of objects to the corresponding classes. The appropriate fuzzy classification model (proper Fuzzy Membership Functions (FMF) or Nearest Neighbor Classification (NNC)) for the assignment of objects into the defined classes was also determined.

2.4 Remote Sensing Methodology

During the stage of problem conceptualization and knowledge acquisition, Remote Sensing methods and techniques were applied for the derivation of useful spectral information concerning the land cover types inherent in the satellite images. These methods included RGB colour composition (Figure 1), PCA analysis, ISODATA unsupervised clustering (Figure 3), decorrelation stretching (Figure 4), as well as texture analysis (textural metrics by Haralick). Appropriate thematic maps derived from this stage were introduced into the knowledge-based system for assisting the discrimination of land cover types and the urban features of interest.

Indicatively, selected thematic maps derived from remote sensing techniques are presented in Figures 3 and 4. These thematic maps were used as input to the knowledge-based system for the identification and discrimination of the land cover classes portrayed in the IKONOS image by exploiting their spectral attributes.

In Figure 3 the thematic map using ISODATA unsupervised clustering of the IKONOS image is presented. Urban features (e.g. roads, squares, parking lots) are illustrated with gray colour. Green areas are presented with hues of green. Buildings or building facilities are depicted with beige and lilac colour. The ISODATA classification map was introduced to the knowledge-based system in order to provide information about the spectral attributes and the spatial context of the land cover classes of the image.



Figure 3: Thematic map using ISODATA unsupervised clustering of the IKONOS image.

In Figure 4, the RGB-421 colour composite derived by decorrelation stretching of the IKONOS image bands is presented. Urban features (e.g. roads, squares, parking lots) are illustrated with bluish-green colour.



Figure 4: RGB-421 colour composite derived by decorrelation stretching of the IKONOS image bands.

2.5 Optimal edge detection algorithms: Implementation

On the band 4 of the IKONOS very high resolution image for the study area of Aghios Stefanos, Attica, the following edge detectors were selected, applied and assessed (Mavrantza and Argialas, 2007):

- The Canny edge detection algorithm (Canny, 1986)
- The Rothwell algorithm (Rothwell *et al.*, 1994)
- The *LOG-LIN* algorithm (Iverson and Zucker, 1995)
- The *SUSAN* operator (Smith and Brady, 1997)
- The anisotropic diffusion algorithm of Black (Black, *et al.*, 1998)
- The Bezdek algorithm (Bezdek, *et al.*, 1998), and
- The *EDISON* algorithm (Meer and Georgescu, 2001)

For each algorithm, the combinations of input parameter sets were selected based on trial-and-error experiments and assessed (a) using the performance evaluation measures of *Pratt* and *Rosenfeld* (Abdou and Pratt, 1979, Kitchen and Rosenfeld, 1981) (Table 1) and (b) by evaluating the optical correspondence to the ground map data for ensuring the “interpretability” of the output edge image.

ALGORITHMS	ROSENFELD METRICS IKONOS - band 4	PRATT METRICS IKONOS - band 4
CANNY	0,6697	0,5290
ROTHWELL	0,6915	0,4811
BLACK	0,6470	0,4530
SUSAN	0,7536	0,5785
IVERSON	0,6372	0,5984
BEZDEK	0,7708	0,4989
EDISON	0,6695	0,5340

Table 1: Performance evaluation metrics (Rosenfeld and Pratt) for the band 4 of the IKONOS satellite image of the study area of Aghios Stefanos

From the stage of the edge detection implementation, the output edge map of the *EDISON* algorithm was selected as the “best” because it fulfils the criteria of (a) good edge extraction (good edge localization, edge connectivity and edge response) and (b) sufficient interpretation ability of the semantic content (according to the qualitative (visual) and quantitative (measurement) criteria set). On the other hand, the output edge map of the application of the Bezdek algorithm was the optimum in terms of visual “interpretability” and good appearance of semantic information (e.g. building boundaries), but needs further post-processing for achieving sufficient edge connectivity and coherence and edge thickness. These two output maps were inserted into the knowledge-based system in order to be segmented and classified for the production of the final urban feature thematic map (Mavrantza and Argialas, 2007).

Using the *EDISON* algorithm, the best output edge map was derived using the following combination of parameters: a) Gradient=4.00, b) Minimum length=7.00, (c)–(e) Non-maxima suppression: Type = line, Rank=0.5 and Confidence=0.5, (f)–(h) T_{high} Hysteresis: Type = line, Rank=0.95 and Confidence =0.98, and (i)–(k) T_{low} Hysteresis: Type = line, Rank =0.99 and Confidence =0.96 (Figure 5 – left image). Using the Bezdek algorithm, the best output edge map was derived using the following input parameter values: $\tau=2.00$ and $Binary_Threshold=60.00$ (Figure 5 – right image)



Figure 5: **Left:** The *EDISON* output edge map. Indicatively, an area with buildings is delineated with a red ellipse. **Right:** The Bezdek output edge map. The same area as in the left image is delineated with a red ellipse on the right image.

2.6 Multi-scale hierarchical segmentation

Multi-resolution hierarchical segmentation proposed by Baatz and Schäpe (1999) and object-oriented image analysis were performed in the environment of the *eCognition* object-oriented image analysis software by Definiens. During this procedure, the input data were segmented into raster primitive regions, which were next assigned into defined thematic land cover classes. The hierarchical segmentation of the input geo-data was performed using five (5) segmentation levels in pre-defined order, data membership factor and selected segmentation parameters, so that the physical boundaries of the object classes of upper and lower levels of segmentation hierarchy won't be affected. In Table 2, detailed information is presented about the input geo-data into the knowledge-based system, their type, the segmentation parameters used as well as the performed segmentation order.

GEODATA TYPE	MULTI-SCALE HIERARCHICAL SEGMENTATION PARAMETERS					FINAL SEGMENTATION LEVEL (Levels 1-5)	SEGMENTATION ORDER (Levels 1-5)
	Scale Parameter	Color Criterion	Shape Criterion	Shape Criterion			
				Compactness	Smoothness		
BEZDEK (membership factor MF=1.0) and EDISON (membership factor 0.5) edge detection maps	10	0.7	0.3	0.5	0.5	5 (upper)	1
Thematic maps of decorrelation stretch (MF=1.0), PCA (MF=1.0), IKONOS (4 bands) (MF=1.0), BEZDEK (membership factor MF=1.0) and EDISON (membership factor 0.5)	6	0.7	0.3	0.4	0.6	4	2
ISODATA map and IKONOS (4 bands) (MF=1.0)	3	0.9	0.1	0.2	0.8	3	3
DSM and IKONOS (4 bands) (MF=1.0)	7	0.6	0.4	0.4	0.6	2	4
EDISON edge detection map (membership factor 1.0)	10	0.7	0.3	0.5	0.5	1 (lower)	5

Table 2: Geo-data type, multi-scale segmentation parameters and performed segmentation order during the segmentation process.

2.7 Knowledge-based identification system of urban features

The knowledge-based identification of the urban and the non-urban features was based on the construction of an object-oriented knowledge base of the inherent land cover classes and the classification (assignment) of the primitive objects derived at the segmentation stage into these semantic classes. A theoretical approach preceded the object-oriented classification process concerning (a) the determination of the proper classes / sub-classes of every segmentation level of hierarchy, (b) the determination of the class attributes based on the spectral, spatial and contextual characteristics of these classes, (c) the determination of the proper fuzzy representation method (Fuzzy Membership Function values or NNC) and finally, (d) the proper classification order at each level of hierarchy which also determines the construction of the fuzzy classification rules.

In Figure 6, the class hierarchy is presented, which was used for the object-oriented identification of the inherent land cover classes (urban and non-urban feature classes).

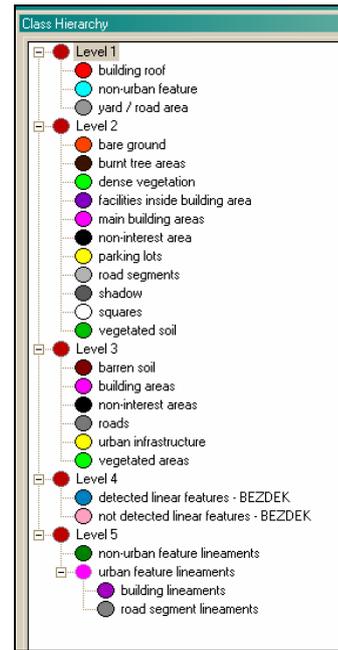


Figure 6: The Class hierarchy used for the object-oriented identification of the inherent land cover classes (urban and non-urban feature classes).

Due to paper length limitation, a fuzzy classification rule (combination of those attributes that determine a class / sub-class) is indicatively presented in Figure 7, while in Figure 8 the corresponding Fuzzy Membership Function of a selected attribute is presented.

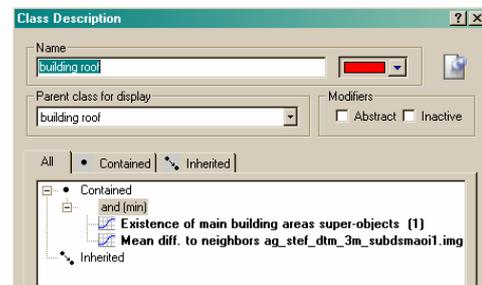


Figure 7: The attribute set of the definition of class **building roof** – Level 1

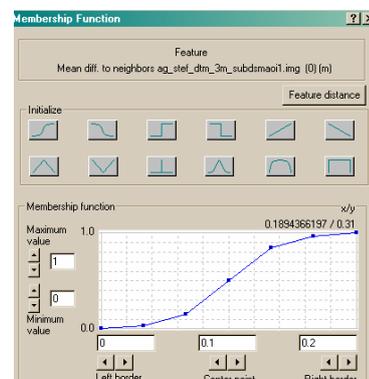


Figure 8: Fuzzy Membership Function for the attribute **Mean diff. To neighbors ag_stef_dsm_subdsmaoi1.img**.

3. DISCUSSION AND CONCLUSIONS

The presented methodological framework was used for the identification of urban features, as well as additional land cover types, contained in the geo-data used. One main aspect from the application of optimal edge detection algorithms was their good performance in terms of coherence, edge localization and high edge response, and therefore they generally provide useful tools towards automated urban / peri-urban mapping by yielding into “interpretable” edge maps. Furthermore, edge maps derived by optimal edge detection can generally be introduced into a knowledge-based system to provide final urban feature thematic maps of useful semantic information, while “isolating” the redundant information of non-interest (non-urban features). The output result of the proposed knowledge-based identification system is expected to be a final map that contains urban features of interest (such as building boundaries and road segments) and additional land-cover classes which consisted the natural scene of the study area (e.g. vegetated regions).

4. REFERENCES

- Abdou, I. E. and W. K. Pratt, 1979. Quantitative Design and Evaluation of Enhancement / Thresholding Edge Detectors. *Proceedings of IEEE*, 67, pp. 753-763.
- Agouris, P., A. Stefanidis and S. Gyftakis, 2001. Differential snakes for change detection in road segments. *Photogrammetric Engineering and Remote Sensing*, 67, pp. 1391-1399.
- Argialas, D., A. Georgopoulos, D. Rokos, Ch. Ioannidis, N. Alvertos, N. Paragios, E. Papadaki, P. Kolokoussis, Ch. Iossifidis, O. Mavrantza, Ch. Papasaika, V. Noutsou, A. Tzotsos, P. Deltisidis, P. Michalis and K. Karantzalos, 2007. Development of advanced Remote Sensing and Photogrammetry Techniques for the implementation of urban geospatial databases, In *Pythagoras: Conference for the Scientific Research at the National Technical University of Athens*, 5-8 July, Plomari, Lesvos
- Baatz, M. and A. Schäpe, 1999. Multiresolution segmentation – an optimization approach for high quality multi-scale image segmentation. In Strobl, Blaschke and Greisebener (Ed.): *Angewandte Geographische Informationsverarbeitung, XI. Beiträge zum AGIT-Symposium*, Salzburg, Karlsruhe, Herbert Wichmann Verlag.
- Bezdek, J. C., R. Chandrasekhan and Y. Attikiouzel, 1998. A geometric approach to edge detection. *IEEE Transactions on Fuzzy Systems*, 6, pp. 52-75.
- Black, M., G. Sapiro, D. Marimont and D. Heeger, 1998. Robust Anisotropic Diffusion. *IEEE Transactions on Image Processing*, 7, pp. 421-432.
- Canny, J. F., 1986. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8, pp. 679-714.
- Heath, M., S. Sarkar, T. Sanocki and K. Bowyer, 1997. A robust visual method for assessing the relative performance of edge detection algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.-PAMI-19, pp. 1338-1359.
- Hofmann, P., 2001. Detecting Informal Settlements from IKONOS Image Data Using Methods of Object Oriented Image Analysis – An Example from Cape Town (South Africa). In: *Jürgens, C. (Ed.): Remote Sensing of Urban Areas / Fernerkundung in urbanen Räumen*. Regensburg: University Regensburg, Germany, pp. 41-42.
- Iverson, L. A. and S. W. Zucker, 1995. Logical / linear operators for image curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17, pp. 982-996.
- Kitchen, L. and A. Rosenfeld, 1981. Edge Evaluation using Local Edge Coherence. *IEEE Transactions on Systems, Man and Cybernetics*, 11, pp. 597-605.
- Mavrantza, O. D. and D. P. Argialas, 2007. Edge Detection Techniques For Extracting Linear Information In An Urban / Peri-Urban Environment”, *Proceedings of the FIG Commission 3 Workshop “Spatial Information Management toward Legalizing Informal Urban Development”*, 28-31 March, Athens, Greece, 15 p.
- Mavrantza, O. D. and D. P. Argialas, 2006. Object-oriented image analysis for the identification of geologic lineaments. *Proceedings of the 1st International Conference on Object-based Image Analysis (OBIA): Automated classification, mapping and updating techniques - Geology, soil, natural resources*, 5-7 July, Salzburg, Austria, 7p, <http://www.commission4.isprs.org/obia06/papers.htm>.
- Mavrantza, O. D. and D. P. Argialas, 2003. Implementation and evaluation of spatial filtering and edge detection techniques for lineament mapping - Case study: Alevrada, Central Greece. *Proceedings of SPIE International Conference on Remote Sensing: Remote Sensing for Environmental Monitoring, GIS Applications, and Geology II*, M. Ehlers (editor), SPIE-4886, pp. 417-428, SPIE Press, Bellingham, WA.
- Meer, P. and B. Georgescu, 2001. Edge detection with embedded confidence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-23, pp. 1351-1365.
- Rothwell, Ch., J. Mundy, B. Hoffman and V. Nguyen, 1994. Driving Vision by Topology. TR-2444 – Programme 4, INRIA, pp. 1-29.
- Smith, S. M. and J. M. Brady, 1997. SUSAN-A New Approach to Low Level Image Processing. *International Journal of Computer Vision*, 23(1), pp. 45-78.
- Ziou, D. and S. Tabbone, 1997. Edge Detection Techniques – An Overview. *TR-195*, Département de Math et Informatique, Université de Sherbrooke, Québec, Canada, pp. 1-41.

ACKNOWLEDGEMENTS

This work was based on data (IKONOS image) and methodology fully conducted under the “EPEAEK-PYTHAGORAS-II” Research Program - National Technical University of Athens, Greece. The “PYTHAGORAS-II” Project is a part of the Operational Program for Educational and Vocational Training II (EPEAEK-II), which is co-funded by the European Social Fund (75%) and National Resources (25%).

POINT POSITIONING ACCURACY OF AIRBORNE LIDAR SYSTEMS: A RIGOROUS ANALYSIS

Nora Csanyi May^{1,2*}, Charles K. Toth²

Department of Civil and Environmental Engineering and Geodetic Science¹
Center for Mapping²
The Ohio State University
1216 Kinnear Road, Columbus OH 43212-1154, E-mail: nora@cfm.ohio-state.edu

Commission I, WG I/2

KEY WORDS: Airborne LiDAR, Accuracy Analysis, Error Propagation, Error Sources, Multi-Sensor Systems, GPS/INS

ABSTRACT:

This paper provides a comprehensive analysis of the achievable point positioning accuracy for airborne LiDAR systems considering all the major potential error sources. Using the law of error propagation, rigorous analytical derivation of error formulas was performed to obtain a reliable assessment of the achievable point positioning accuracy. For practical use, based on the derived analytical formulas, accuracy figures and tables have been developed. In this paper select example figures illustrate the achievable point positioning accuracies of state-of-the-art LiDAR systems for various accuracies of the parameters that influence the LiDAR point positioning accuracy. The analytical derivation-based accuracy plots can also be used as a tool for choosing the right system or system configuration for a desired mapping accuracy, and to help the flight planning, i.e. selecting optimal flight parameters for a given system to achieve the desired point positioning accuracy. In addition, the developed error formulas can also facilitate the analysis of the effects of individual error sources on point positioning accuracy, although due to the size limitations of the paper this analysis is not included in this paper.

1. INTRODUCTION

LiDAR systems are complex, multi-sensor systems consisting of at least three sensors, the GPS and INS navigation sensors, and the laser scanner system. Consequently, proper system calibration, including individual sensor calibration, inter-sensor calibration, and time synchronization between system components is crucial in achieving the required mapping accuracy. Furthermore, besides errors in the calibration parameters, there are several other error sources that can degrade the accuracy of the derived ground coordinates, such as, for example, errors in the navigation solution (position and attitude errors), range measurement errors, etc. In addition, the effect of the various errors is influenced by the various flight parameters (flying height, flying speed, etc.), terrain characteristics, and system settings, and accordingly, the dependency of point positioning accuracy on the various error sources is very complex. As a consequence, the reliable accuracy assessment and performance validation of the derived mapping products is a very challenging task.

Most publications, discussing the effects of different error sources on the point positioning accuracy of LiDAR systems typically focus on a single or a few error sources and do not discuss the combined effect of all error sources. Baltsavias (1999), for example, provides an overview of basic relations and error formulas concerning airborne laser scanning. Schenk (2001) provides a summary of the major error sources for airborne laser scanners and error formulas focusing on the effect of some systematic errors on point positioning. A number of papers empirically evaluates the achieved accuracy of specific

mapping projects, normally using ground control as reference (Latypov, 2002; Hodgson and Bresnahan, 2004; Hodgson et. al., 2005; Peng and Shih, 2006). LiDAR vendors provide specifications on the approximate accuracy that can be expected from their systems. These values, however, are mostly valid under specific circumstances (for specific flying height, GPS baseline, etc.), and only consider a few error sources, and consequently are frequently either too optimistic or too pessimistic. Furthermore, some of the vendors do not clearly state what error sources are considered when they provide the accuracy specifications, which makes it difficult and, in some cases, nearly impossible to compare the achievable accuracies of different systems from different vendors. For example, some LiDAR vendors specify the achievable point accuracy considering the GPS errors, while others do not include this error in their accuracy specifications. In conclusion, no generally accepted, comprehensive and reliable accuracy assessment tool exists to help with flight or project planning in order to achieve the desired accuracy of the final product of LiDAR systems.

This paper is intended to fill the void by providing a comprehensive accuracy assessment tool for airborne LiDAR systems that considers all the major potential error sources, and consequently a reliable assessment of the achievable point positioning accuracy can be obtained. The first section of the paper provides an overview of the analytical derivation of the error formulas. In the second section, example plots of the achievable point positioning accuracies of different systems for various accuracies of the parameters that influence the point positioning accuracy are shown and analyzed.

* Corresponding author.

2. RIGOROUS ERROR PROPAGATION

To reliably determine the achievable point positioning accuracy of airborne LiDAR systems, all the major potential error sources that influence the point positioning accuracy have to be considered in the analytical derivations. The random errors listed in Table 1 were considered during the analytical derivations. The 2nd column of Table 1 also shows the symbols for the respective standard deviation values as used in this paper; the laser beam divergence (γ) itself is obviously not an error term, however, its effect, the finite footprint size represents an additional random error source.

Error source	Used symbol
Navigation solution errors	
Position errors	$\sigma_X, \sigma_Y, \sigma_Z$
Attitude angle errors	$\sigma_\omega, \sigma_\phi, \sigma_\kappa$
Errors in the determined boresight misalignment angles	$\sigma_{\omega_b}, \sigma_{\phi_b}, \sigma_{\kappa_b}$
Range measurement error	σ_r
Scan angle error	σ_β
Laser beam divergence	γ

Table 1. Considered error sources

The error formulas for point positioning accuracy were derived based on the LiDAR equation (Eq.1.) via rigorous error propagation, see also Figure 1.

$$r_M = r_{M,INS} + R_{INS}^M (R_L^{INS} \cdot r_L + b_{INS}) \quad (1)$$

where

r_M	—	3D coordinates of object point in the mapping frame
$r_{M,INS}$	—	Time dependent 3D INS coordinates in the mapping frame, provided by GPS/INS (refers to the origin of the INS body frame)
R_{INS}^M	—	Time dependent rotation matrix between the INS body and mapping frame, measured by INS
R_L^{INS}	—	Boresight matrix between the laser frame and INS body frame
r_L	—	3D object coordinates in laser frame
b_{INS}	—	Boresight offset vector

Using the law of error propagation based on the covariance matrix of the error sources listed in Table 1, the covariance matrix of the 3D LiDAR positions was derived as described in Eq. 2.

All analytical derivations were implemented in MATLAB environment; due to size limitations of this paper, the derived formulas are not shown here. The effect of laser beam divergence was considered separately from the error propagation (but later combined) since the horizontal error due to the footprint size can be characterized by uniform distribution instead of the normal distribution. Furthermore, the boresight offset component was assumed to be error free since its effect is negligible, as compared to the effects of other errors.

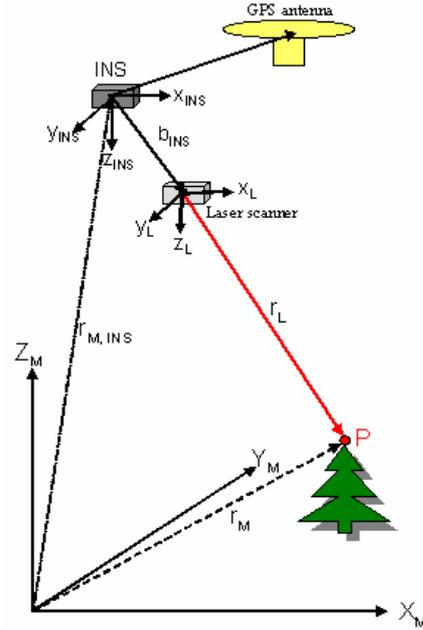


Figure 1. LiDAR system components and definitions.

$$C_{LiDAR} = ACA^T \quad (2)$$

where

C_{LiDAR} [3x3]	—	Covariance matrix of the LiDAR point coordinates
C [11x11]	—	Covariance matrix of the INS position, INS attitude angles, boresight angles, measured range, and the scan angle
A [3x11]	—	Jacobian matrix containing the partial derivatives of X,Y, Z LiDAR coordinates with respect to the different random variables in the LiDAR equation

The following assumptions were made during the analytical derivations: Flat terrain was assumed, sloped terrain will cause additional errors in the vertical coordinates as compared to the error propagation results; however, this effect can easily be considered and accounted for separately. The error formulas were developed considering that the scanning is performed in a vertical plane perpendicular to the flight direction. The range measurement accuracy is assumed to be independent of the flying height and scan angle, which for earlier systems was not exactly the case, since longer range due to higher flying height and larger scan angle meant weaker signal response, and consequently less accurate range measurement. However, according to LiDAR vendors, for the state-of-the-art systems, the range measurement accuracy does not noticeably degrade for longer ranges. During the derivations the various errors mentioned above were considered to be uncorrelated with each other, however, any correlation between variables could easily be considered by changing the covariance matrix to non-diagonal.

Using the derived error formulas, based on the accuracy of the parameters listed in Table 1, the achievable point positioning accuracy can be computed for any given LiDAR system operated at different flying heights, etc.

3. EXAMPLES OF ACHIAVABLE LIDAR POINT POSITIONING ACCURACY

In the following examples, select accuracy plots are shown. The figures were generated using typical accuracy values for the various error sources that represent the state-of-the-art LiDAR systems. For the flight parameters, such as flying height, and the maximum scan angle values – or in some figures value ranges – that are frequently used in LiDAR mapping, were considered.

The accuracy of the navigation parameters, the position and attitude angle accuracies used in these examples are chosen based on the post-processed accuracy specifications of the Applanix POS/AV™ systems. Since the examples are intended to illustrate the performance that can be expected from state-of-the-art systems, only the accuracies provided by the high-end systems, such as POS/AV™ 410-610, (www.applanix.com) were considered for the examples. Consequently, for the generation of the following accuracy plots, the accuracy ranges listed in Table 2 were used. For the accuracy of the boresight misalignment angles, typical achievable standard deviations of the calibrated boresight misalignment angles were considered (Burman, 2001; Skaloud and Lichti, 2006). Table 2 contains the values that were used in the generated examples. The range measurement was assumed to have a 1 cm standard deviation (1σ), this value is based on the system specifications of state-of-the-art LiDAR systems. It should be emphasized that our example plots assume hard surfaces; the ranging accuracy used in our computations is obviously not valid in vegetated areas. The accuracy of scan angle measurement is typically not addressed in the literature or in the system specifications provided by LiDAR vendors. In the examples below, a quantization error with 0.0007° standard deviation (1σ) was assumed, which was mentioned as a valid value for the Riegel LMQ-280 system in (Campbell, et. al, 2003). The laser beam divergence (γ) of 0.3 mrad was considered based on LiDAR system specifications of modern LiDAR systems.

Parameter	Value (1σ)
σ_X, σ_Y	5-15 cm,
σ_Z	7.5-22.5 cm ($1.5*\sigma_X, \sigma_X=\sigma_Y$)
$\sigma_\omega, \sigma_\phi$	10''-30''
σ_κ	20''-60'' ($2*\sigma_\omega, \sigma_\omega=\sigma_\phi$)
$\sigma_{\omega_b}, \sigma_{\phi_b}$	10''
σ_{κ_b}	30''
σ_r	1 cm
σ_β	0.0007°

Table 2. Standard deviation values of parameters assumed for the illustrated examples

For the sake of simplicity, the plots below illustrate the accuracy of the LiDAR point coordinates in a local right-handed coordinate system that has its X-axis aligned in the flight direction, Y-axis points to the scan direction, and Z up. Furthermore, for the generation of these plots all three aircraft attitude angles (roll, pitch, and heading) were assumed to be zero for the same reason, but any other value could be used in the derived accuracy formulas. In all plots shown in the paper the vertical (Z) accuracy is shown in red, the accuracy in the scan direction (Y) is marked with green color, and in the flying direction (X) it is shown in blue.

3.1 The Effect of Attitude Angle Errors

Figure 2 and 3 illustrate the effect of attitude angle errors on the achievable point positioning accuracy as a function of scan angle for 600 m and 1500m flying height, respectively. To better show the effects of the attitude angle errors, in these figures all other variables are considered to be error free. The ‘Sigma omega phi’-axis of these figures show the $\sigma_\omega, \sigma_\phi$ values, the σ_κ value was taken according to the ratio shown in Table 2. As the figures show, the attitude angle errors have stronger effect on the horizontal positions than on the vertical one, and errors in all three coordinate directions increase for higher flying heights. Furthermore, the effect of attitude angle errors on the coordinate accuracy in the scan direction does not change with the scan angle, while the accuracy in both the flying direction and the vertical coordinate direction degrades with increasing scan angles (towards the sides of the LiDAR strips). This increasing effect of attitude angle errors in the flying direction is caused by the κ angle error that has increasing effect for larger scan angles, while the accuracy degradation of the vertical coordinates towards the sides of the strip is due to the ω attitude angle error that also has increasing effect with larger scan angles.

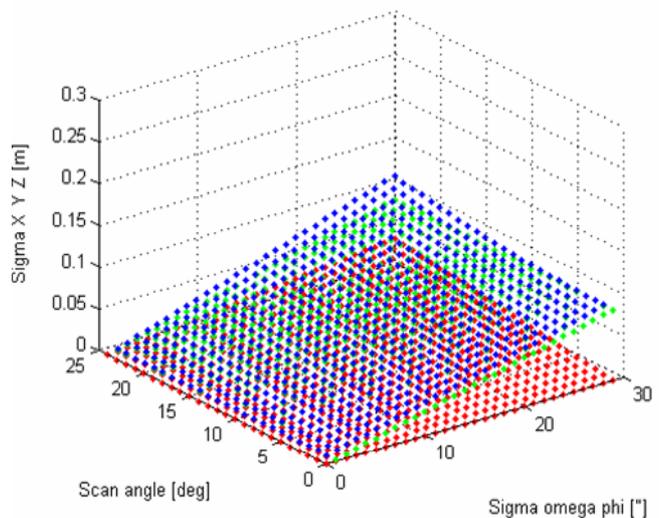


Figure 2. Effect of attitude angle errors on point positioning as a function of scan angle for H=600 m.

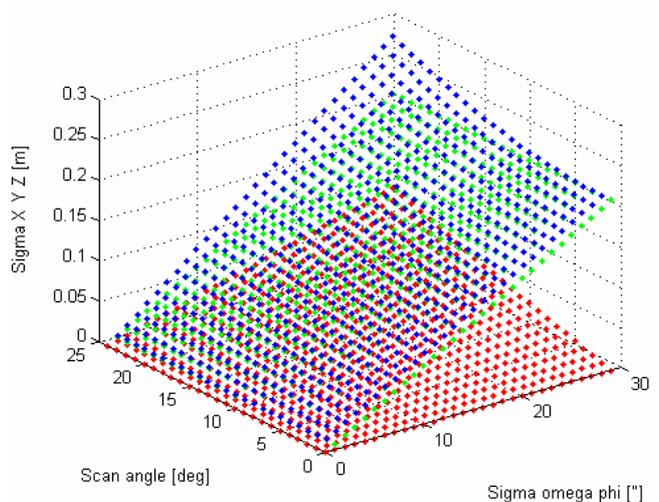


Figure 3. Effect of attitude angle errors on point positioning as a function of scan angle for H=1500 m.

3.2 The Effect of All Navigation Errors

To illustrate the effect of the accuracy of all the navigation parameters, Figure 4 and 5 illustrate the achievable point positioning accuracy as a function of the accuracy of the aircraft position and attitude angles for 600 m and 1500m flying height, respectively. To better show the effects of navigation errors, in these figures all other variables are considered to be error free. The figures illustrate the point positioning accuracies at 10° scan angle, and in order to also show the effect of aircraft position and the attitude errors separately from each other; the point positioning accuracies were computed starting at zero navigation errors, but the realistic values are in the ranges shown in Table 2. As Figure 4 illustrates, for lower flying heights, the accuracy of the aircraft position has a relatively larger effect on the point positioning accuracy as compared to the angular errors (in particular, on the vertical coordinates). Note that the higher positioning errors in the vertical coordinates for larger aircraft position errors in this figure are caused by the 1.5 ratio of $\sigma_z/\sigma_{X,Y}$ (which is rather realistic) used in the computation. This might be a bit surprising, since LiDAR vertical accuracy is known to be better than the horizontal one, which, as the other figures show, is normally true due to the other errors that affect the horizontal position more (especially the beam divergence and attitude errors). As Figure 5 shows, as the flying height increases, the attitude angle errors have increasing effect on the point positioning accuracy, while the effect of aircraft position errors do not increase.

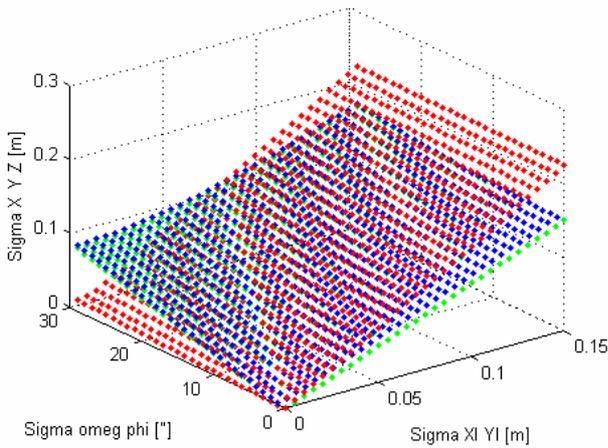


Figure 4. Effect of navigation errors on point positioning for H=600 m at 10° scan angle.

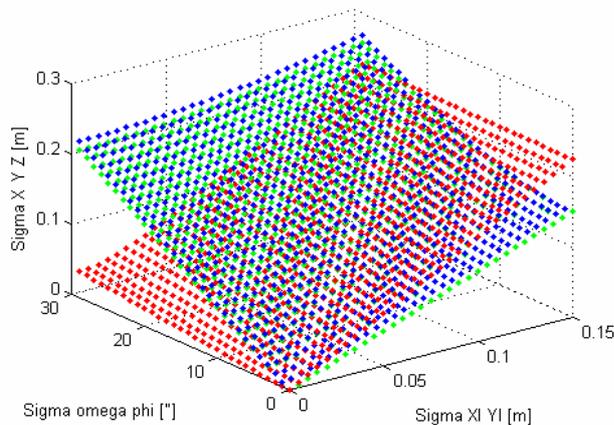


Figure 5. Effect of navigation errors on point positioning for H=1500 m at 10° scan angle.

3.3 All Error Sources Considered

Figure 6 and 7 illustrate the point positioning accuracies for the same cases as Figure 4 and 5, but in these figures all other error sources with accuracies listed in Table 2 were also considered. The zero aircraft position error and zero attitude error (which is obviously not a realistic case) is only intended to show the effect purely of all other error sources excluding the navigation errors. As these plots show, these other error sources have stronger effect on the horizontal position than on the vertical, and especially for higher flying height, the vertical LiDAR point accuracy is indeed better as compared to the horizontal accuracy (except for some unrealistic cases when the accuracy of the aircraft position is much worse than that of the attitude angles).

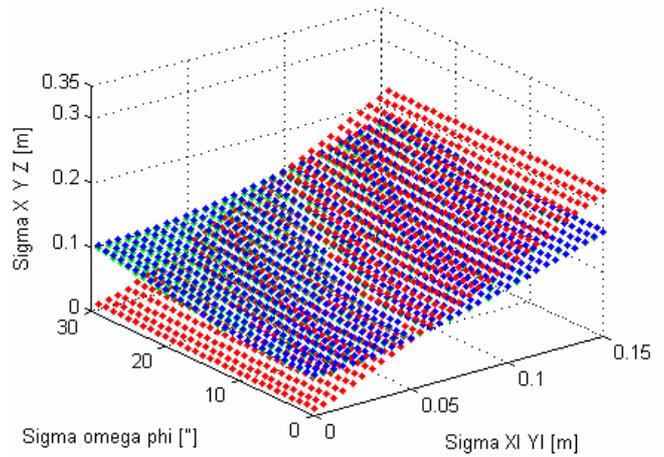


Figure 6. Standard deviation of point positioning for H=600m at 10° scan angle as a function of navigation errors, all errors considered.

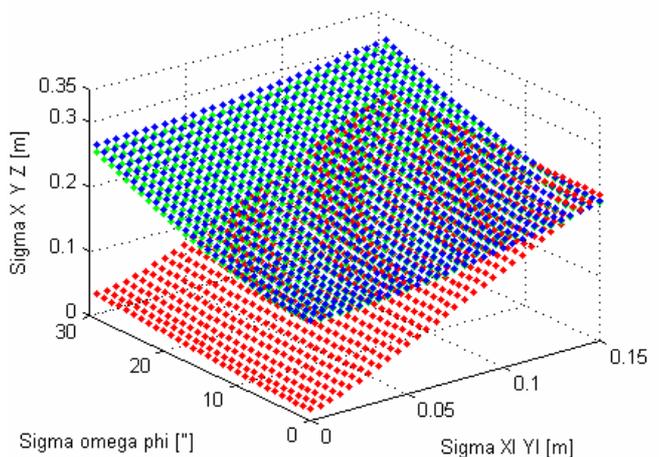


Figure 7. Standard deviation of point positioning for H=1500m at 10° scan angle as a function of navigation errors, all errors considered.

3.4 Flying Height and Scan Angle Dependency

Finally, Figure 8 illustrates the point positioning accuracies achievable as a function of flying height and scan angle. This plot is intended to show the case of a state-of-the-art LiDAR system including a highly accurate navigation solution. For this figure the following standard deviations were considered: $\sigma_x=\sigma_y=5\text{cm}$, $\sigma_z=7.5\text{cm}$, $\sigma_\omega=\sigma_\phi=15''$, $\sigma_\kappa=30''$, the accuracy of the other parameters were assumed as shown in Table 2. As

the figure illustrates, as the flying height increases, the accuracy of the vertical coordinates do not significantly decrease (especially for smaller scan angles), while the horizontal point positioning accuracy does. Towards the LiDAR strip edges (with higher scan angles) all three coordinates show degrading accuracy; in the scan direction this degradation is small, while in the flying direction and vertical direction the errors increase more. The higher degradation in accuracy in the flying direction as compared to the scan direction – that is noticeable in the figure – can be explained by the fact that errors in the heading and in the κ boresight angle affect the accuracy in the flying direction, and this effect significantly increases with higher scan angles.

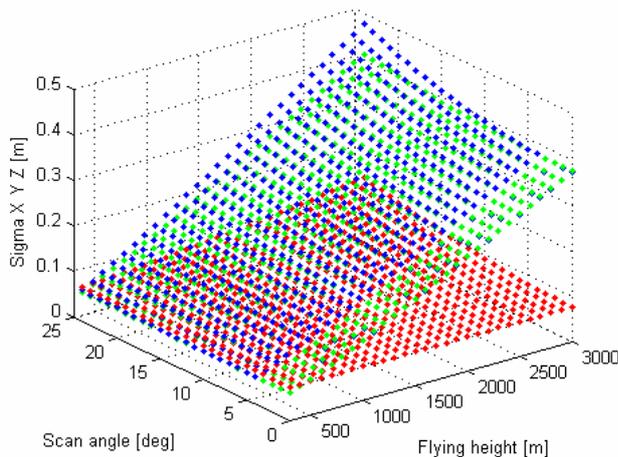


Figure 8. Standard deviation of point positioning as a function of flying height and scan angle, all errors considered.

4. CONCLUSIONS

This paper presented a comprehensive analysis of the achievable point positioning accuracy of airborne LiDAR systems using rigorous analytical derivations via error propagation. LiDAR systems in reality are rather complicated and consequently not all potential error sources could be accounted for in this analysis. However, as all the major error sources were considered, therefore, we believe that for well-calibrated LiDAR systems the derived formulas provide a rather realistic and reliable accuracy assessment when the accuracies of the considered parameters that influence the point positioning accuracy are reasonably well known.

Using the derived error formulas, based on the accuracy of the navigation solution, the boresight misalignment angles, the ranging and scan angle accuracy, and laser beam divergence, the achievable point positioning accuracy can be computed for any given LiDAR system, operated at different flying heights, etc. The accuracy plots that were derived based on the analytical derivations can be used as a tool for choosing the right system for given application requirements, and to help with flight planning to decide on the optimal flying height, and maximum scan angle to achieve the desired point positioning accuracy. Due to the length limitations of this paper, the effects of individual error sources on point positioning accuracy were not analyzed here, but the developed error formulas also facilitate that type of analysis.

5. REFERENCES

- Baltsavias, E.P., 1999. Airborne laser scanning: basic relations and formulas. *ISPRS Journal of Photogrammetry & Remote Sensing* Vol. 54, pp. 199-214.
- Burman, H., 2000b. Calibration and Orientation of Airborne Image and Laser Scanner Data Using GPS and INS, PhD dissertation, Photogrammetry Reports No. 69, Royal Institute of Technology, Stockholm, 107 p.
- Campbell, J. L., M. Uijt de Haag, and A. Vadlamani, 2003. The Application of LiDAR to Synthetic Vision System Integrity, *IEEE Digital Avionics Systems Conference, 2003, DASC '03*, 12-16 Oct. 2003, Vol. 2., pp. 91-97.
- Hodgson, M. E. and P. Bresnahan, 2004. Accuracy of Airborne LiDAR-Derived Elevation: Empirical Assessment and Error Budget, *Photogrammetric Engineering & Remote Sensing*, Vol. 70, No. 3., March 2004, pp. 331-339.
- Hodgson, M.E., J. Jensen, G. Raber, J. Tullis, B. A. Davis, G. Thompson, and K. Schuckman, 2005. An Evaluation of Lidar-derived Elevation and Terrain Slope in Leaf-off Conditions, *Photogrammetric Engineering & Remote Sensing*, Vol. 71, No. 7., July 2005, pp. 817-823.
- Latypov, D., 2002. Estimating relative lidar accuracy information from overlapping flight lines. *ISPRS Journal of Photogrammetry and Remote Sensing*, Volume 56, Issue 4, July 2002, pp. 236-245.
- Peng, M. H. and T. Y. Shih, 2006. Error Assessment in Two Lidar-derived Datasets. *Photogrammetric Engineering & Remote Sensing*, Vol. 72, No. 8, August 2006, pp. 933-947.
- Schenk, T., 2001. Modeling and Analyzing Systematic Errors in Airborne Laser Scanners. *Technical Notes in Photogrammetry*, vol. 19. The Ohio State University, Columbus, USA. 46 p.
- Skaloud, J. and D. Lichti, 2006. Rigorous approach to bore-sight self-calibration in airborne laser scanning. *ISPRS Journal of Photogrammetry & Remote Sensing* Vol. 61, pp. 47-59.

NATIONAL AND REGIONAL SCALE DEMS CREATED FROM AIRBORNE INSAR

Bryan Mercer

Intermap Technologies Corp., 1000, 736 – 8th Ave. S.W., Calgary, AB, Canada, T2P 1H4
bmercer@intermap.com

Commission I, WG I/2

KEY WORDS: DEM, DSM, DTM, InSAR, Interferometric SAR, Mapping

ABSTRACT:

The objective of this paper is to provide an update on a program to create a single uniform DEM (Digital Elevation Model) for Western Europe and for the USA over the next two years. This program is referred to as NEXTMap[®] and utilizes airborne InSAR (Interferometric Synthetic Aperture Radar) technology. In this paper the technology and the platforms with which this program is to be accomplished will be summarized. Some details on the operational implementation and current status will be provided. Moreover it will be placed in context with other programs such as the SRTM, the planned Tandem-X mission and with other technologies such as airborne lidar.

1. INTRODUCTION

The use of DEMs is widely spread and growing, not only in the traditional mapping world but increasingly in support of new applications that are driven by consumer interests. In this new environment, not only do required levels of detail and implicit accuracy vary according to application, but price and current availability are major considerations for the user, many of whom come from outside the geomatics industry. An additional consideration is that some applications, in order to be effective, transcend local political boundaries and require uniform data-sets across regional, national and even continental scales. Meanwhile the advances of enabling technologies such as GPS, communications bandwidth, storage capacity and processing power have been instrumental in the growth of both numbers and capability of systems for DEM creation including both passive and active systems. Among the active systems, both lidar and InSAR (Interferometric SAR) have become major sources of three-dimensional information.

At the high performance end of the performance scale, modern lidars with point densities of several points per msq can achieve DEM accuracies better than 30 cm RMSE routinely, with 2-3 times better accuracy given sufficient effort. Typically prices range upwards from US\$100/kmsq for lidar-generated DEMs and the project areas are often local (some exceptions exist in both the USA and Europe). At the other end of the performance/cost/availability scale, the SRTM DEM with coverage over approximately +/- 60 degrees latitude on a 3" grid (about 90 meters at the equator) and vertical accuracy generally in the 4 -12 m RMSE range is almost universally available at very little cost.

The subject of this paper is Intermap's NEXTMap program in which airborne InSAR is used to create DEM (and image) products with vertical accuracy of 1 meter, sample spacing's of 5 meters, and national-scale coverage in Western Europe and the USA (excluding Alaska) by 2009. In the following sections we will provide a brief background with respect to the InSAR technology, and summarize the specifications and validation of the various DEM and image products. This will be followed by

a discussion of the NEXTMap concept with an update of the current implementation status and a description of the current capacity of the acquisition and processing elements required to achieve the NEXTMap goals and schedule. Examples of the DSM products will be provided in the context of some of the new application opportunities that are developing. Finally the competitive environment relative to alternative technologies will be addressed, followed by a summary and conclusions.

2. BACKGROUND

2.1 InSAR Summary

The interferometric process has been widely discussed in the literature, (e.g. Zebkor and Villsenor (1992), Bamler and Hartl (1998), Rodriguez and Martin (1992)). The geometry relevant to height extraction, 'h', is illustrated in Figure 1. If the two antennas, separated by baseline 'B', receive the back-scattered signal from the same ground pixel, there will be a path-difference 'δ' between the two received wave-fronts. The baseline angle 'θ_b' is obtainable from the aircraft inertial system, the aircraft height is known from differential GPS and the distance from antenna to pixel is the radar slant range. Then it is simple trigonometry to compute the target height 'h' in terms of these quantities as shown in equations 1-3.

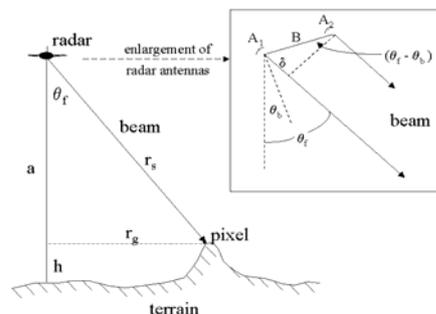


Figure 1. Schematic of Airborne IFSAR Geometry.

$$\sin(\theta_r - \theta_b) = \delta/B \quad (1)$$

$$\delta/\lambda = \varphi/(2*\pi) + n \quad (2)$$

$$h = H - r_s \cos(\theta_r) \quad (3)$$

The path-difference ‘ δ ’ is measured indirectly from the phase difference ‘ φ ’ between the received wave fronts (eqn. 2). Because the phase difference ‘ φ ’ can only be measured between 0 and 2π (modulo 2π), there is an absolute phase ambiguity (‘ n ’ wavelengths) which is normally resolved with the aid of relatively coarse ground control. A “phase unwrapping” technique completes the solution. Thus the extraction of elevation is performed on the “unwrapped” phase. Often the IFSAR is operated in a so-called ping-pong mode which effectively doubles the value of the geometric baseline B . These equations become the basis for sensitivity and error analysis (e.g. Rodriguez and Martin (1996)). For ‘single-pass’ InSAR airborne systems as described in this work, the signals are received almost simultaneously so that errors induced by temporal-decorrelation are not a factor as is the case for satellite systems such as ERS and Radarsat which operate in a ‘repeat-pass’ mode. Provided the baseline length, position (from DGPS) and attitude (from coupled GPS/inertial) are adequately controlled and/or measured, the dominant noise-like error source arising out of these sensitivity equations is ‘phase noise σ_φ ’ so that the signal-to-noise ratio, which is a function of flying height among other system-related factors, becomes a means of (partly) controlling height error specifications. That is, other parameters being fixed, the height noise will increase as a function of flying height. For example, DEMs created from the STAR-3i system (Table 1), when operated at about 9km altitude, has a height-noise level of about 0.5 m (1 sigma, 5 m sample spacing) at the far edge of the swath. Systematic errors, with reference to STAR-3i DEMs, are usually slowly-varying and arise from a variety of sources but are limited through calibration, operational and processing procedures

The schematic idealization of Figure 1 is replaced of course by many factors in the practical implementation of InSAR. For example a complex image containing phase and magnitude information is created from the signal received at each antenna. Subsequent operations on the complex images allow three ortho-rectified products to be derived: DEM, Magnitude and Correlation. The DEM, as noted earlier, is usually referred to as a DSM in recognition that the received signal relates to the scattering surface which may be the terrain or could be an object upon the terrain, natural or otherwise. The magnitude is often referred to simply as an ORI (Ortho-Rectified Image). In relatively open urban or forest situations, it is possible to create a DTM (Digital Terrain Model) from the DSM (Wang, et. al. (2001)) and this has been offered in past as a Core Product along with the DSM and ORI (see Table 2 below). More recent alternatives to this process are beyond the scope of this paper.

2.2 Current Intermap Airborne InSAR Systems

Intermap has developed and currently operates five airborne InSAR systems (Table 1). The first of these systems STAR-3i[®], has been operated by Intermap since 1996. STAR-3i, is an X-band, HH polarization InSAR flown on a Learjet 36 (Tennant and Coyne, 1999). In the last few years, Intermap has replaced all of the software and most of the hardware in order to improve product quality and efficiency of operation. In particular this has led to higher resolution images, and better vertical accuracy of the DEMs (Table 2). Equally important,

software automation has led to improvements in processing throughput. The net result of the modifications is higher quality data sets moving toward decreasing costs and wider availability. The second system, named TopoSAR[®] (formerly called AeS-1 [Schwäbisch and Moreira, 1999]), and is flown on an Aerocommander turbo-prop platform. In addition to X-Band, HH single-pass IFSAR, it supports repeat-pass, fully polarized P-Band InSAR and is currently being used mainly as a research platform.

In order to increase production capacity, a new InSAR system architecture has been created, incorporating elements of STAR-3i and TopoSAR design. The third InSAR system, created in 2004, is named STAR-4, is based completely on the new architecture and is flown in a King Air prop-jet. Two additional systems based on this architecture have subsequently been developed by Intermap in 2006. At the time of this preparation, STAR-5, which is also flown on a King Air platform has been operational for 7 months, while STAR-6, integrated into another Lear Jet, has been operational for about 3 months. Current specifications of the four STAR systems used in the NEXTMap program are summarized in Table 1 and the platforms shown in Figure 2. In principle the acquisition rate for a single system could reach about 7,000 km²/hr for the Lear Jets. In practise, however, the effective acquisition rate is considerably reduced by turning times at end-of-line, side lap (terrain-dependent) and the requirement for partial re-flights due to factors which would otherwise drive the data outside specification.

System / Parameter	STAR-3i	STAR-4	STAR-5	STAR-6
Operational year (Initial)	2002 (1996)	2004	2007	2007
Platform	Lear 36	King Air 200T	King Air 200T	Lear 36
Typical Speed (km/hr)	720	430	430	720
Wavelength / Polarization	3 cm / HH	3 cm / HH	3 cm / HH	3 cm / HH
Peak Transmit Power (kW)	8	8	8	8
Bandwidth (MHz)	135	135 / 270	135 / 270	135 / 270
Baseline Length (m)	0.92	1.02	1.02	1.04
Polarization	HH	HH	HH	HH
Look Direction	Left or Right	Left or Right	Left or Right	Left or Right
Platform Altitude (km)	6 – 12	4 – 8.5	6 – 12	4 – 8.5
Swath Width (km)	8 – 15	6 – 11	8 – 15	6 – 11

Table 1: Selected System Parameters of Intermap’s Airborne InSAR Systems (STAR Series only)



Figure 2. Clockwise from upper left: STAR-3i, STAR-4, STAR-6 and STAR-5.

2.3 Product Specifications

The core products available from Intermap’s online store include an Ortho-rectified Radar Image (ORI) a Digital Surface Model (DSM) and the bare earth Digital Terrain Model (DTM). X-band images are at 1.25-m resolution with similar horizontal accuracy. DSM and DTM are posted at 5m spacing. The elevation products are available in three standard vertical accuracy specifications as illustrated in Table 2 below. It is worth noting that all four of the STAR family of sensors are able to achieve these product specifications despite the nuance of individual system design or platform specifics. Apart from these CORE specifications, other accuracies and image/DEM resolution can be supported to meet specific requirements. Optical/radar merged products are now also becoming available. Varying flying altitudes and operating modes, enables different accuracy specifications to be achieved which may be reflected in cost.

Product Type	DSM		DTM	
	RMSE	Spacing	RMSE	Spacing
I	0.5	5	0.5	5
II	1	5	1	5
III	3	10	-	-

Table 2: Intermap Core Product specifications for InSAR DSMs and DTMs. All units are meters. RMSE refers to vertical accuracy and is with respect to terrain that is moderately sloped, bare (DSM) and unobstructed. DTM specifications apply to areas for which the forest or other above ground cover is ‘patchy’ to a maximum scale of about 100 meters. Details of these specifications may be found at www.intermaptechnologies.com.

2.4 Operational Flow

The operational flow consists of four major stages: (1) planning and acquisition, (2) interferometric processing, (3) editing and finishing, and (4) IQC (Independent Quality Control) after which the data are delivered to the data base repository (the ‘i-Store’). The operational concept has evolved to accommodate the requirements imposed by the current NEXTMap® goals as well as custom projects. The NEXTMap® Europe and USA objectives alone require the data acquisition for an area incorporating more than 10 million km² by the end of 2008. All aspects of production are managed with rigorous QC checks throughout and within the framework of the companies’ ISO9000 certification.

2.4.1 Planning and Acquisition: Typically blocks of 200 km x 200 km are planned. Within these blocks, parallel flight lines are planned according to terrain to ensure that there will be no gaps between swaths. Three tie lines per block are flown, orthogonal to the main data passes in order to assist in the removal of systematic errors during the processing stage. Radar reflectors are pre-sited at the ends of the tie lines and surveyed in precisely. These ultimately fix the DEM and ORI to the reference ellipsoid. In order to achieve the Type II DSM vertical accuracy specification, GPS baselines are kept within about 75km. Formerly this meant deploying ground crews to operate GPS ground stations in the acquisition area: however the densification of CORS networks in the USA and similar GPS networks in Europe, is easing this requirement. Moreover, a recent improvement in process and software has enabled longer line lengths, to 1200 km, to be flown by the Lear Jets, thereby reducing the fraction of lost time due to aircraft turns at the end of lines. Navigation and certain other data are processed and analysed after each mission in order to determine whether the quality limits have been satisfied for the particular data type specified. Data which do not satisfy the quality criteria must be re-acquired, normally before the aircraft has departed from the area.

2.4.2 Interferometric Processing: Because of the multiple platforms, high data volumes, differing product specifications, varying schedule priorities, and the demand for high throughput, the processing task is very complex. This part of the process, now performed in a single processing centre, has been automated to a considerable extent with internally developed software and procedures. The system is PC-based (64-bit blade processors). In particular, the processing architecture has been developed to be scalable in order to accommodate the increasing data volume as acquisition efficiency and platform numbers have increased. The process ingests the raw signal data (‘phase history’), ancillary data and navigation data, performs the interferometric processing to ultimately create strips of image and elevation data. Systematic errors are reduced through an internally-developed adjustment process using the controlled orthogonal tie lines that run across the data block. Output products in the form of 7.5’x7.5’ tiles are created for subsequent refinement.

2.4.3 Editing and Finishing: An independent system referred to as IES (Intermap Editing System) has been developed, incorporating a set of editing tools within a stereo-viewing environment. The input data tiles (images and DSMs) are ortho-rectified so false stereo is created to enable the image to be viewed and manipulated in stereo. The human editor uses the tool set to address hydrological features (e.g. flatten water bodies, ensure rivers run down hill), edit transportation features (roads, railways, airports, etc) and to edit radar artefacts. A set of editing rules determines a consistent application of the IES tools across different geography and across different editors. Editing is the most human intensive component of the operational activity and incorporates many quality control steps within the process. At the end of the process, 15 km x 15km tiles of DSM, DTM and ORI are available for IQC.

2.4.4 IQC (Independent Quality Control): A group separate from the production team reviews the IES output tiles for editing completeness and accuracy. The accuracy checks are usually with respect to publicly-available GCP data such as the NGS (National Geodetic Survey) network in the USA. Tiles that are not judged complete at the 95% level, or whose accuracy checks are outside specification are returned to the production side for re-processing as required. A summary accuracy report for the state of California is noted in the caption to Figure 3. The (normally) successful tiles are forwarded to the i-Store at this stage for entry into the data-base. Subsequent to this stage the data are available to the public under license. A global interface tool allows regions of interest to be displayed for availability and detail including thumbnail presentations.

3. NATIONAL MAPPING PROGRAMS: NEXTMAP

NEXTMap is the term used by Intermap to describe its InSAR-based national and regional mapping programs. Specifically the concept is to make DSM, DTM and ORI products generally available in a seamless fashion over national and trans-national regions where multiple applications and markets may benefit. By retaining ownership and licensing the data to multiple users, the cost is shared, making it feasible for public and private organizations to have access to these data sets or to parts of them. The Type II specification for the DSM creates a level of detail (1m RMSE vertical accuracy, 5 meter samples) intermediate between lidar or photogrammetrically-produced products on the one hand, and SRTM or SPOT5 products on the other (Table 3). The associated ORI carries a resolution of 1.25 m and horizontal accuracy less than 2m RMSE.

NEXTMap Britain was implemented in 2002/2003 (England and Wales) and subsequently extended to include Scotland (for a description, see Mercer, 2004). On the basis of the success of that project, as well as lessons learned, the decision was made to proceed with a NEXTMapUSA project with the goal of 2008 completion. As of April, 2007 about 1/3rd of the 8 million km² had been acquired and over 1/4 of these data had been interferometrically processed. This rate is rapidly accelerating because of the introduction into operations of STAR-6 since then. An example is shown in Figure 3 of the DSM of the State of California.



Figure 3: NEXTMAP USA Example - California DSM
Validated vertical accuracy (1430 check points) 0.76 m RMSE

3.1 NEXTMap Europe: Status

Twelve countries in western Europe are currently scheduled in the NEXTMap Europe program (Figure 4) and this list will

likely be expanded. The data acquisition phase for Germany was completed in a 3-month period of 2006 and the completed products should become available in September 2007. Forecasts for product availability of the remaining countries is shown in Figure 4. Two factors will likely impact these forecasts: (1) the time it takes to obtain overflight permits vary from country to country, and (2) priorities are subject to market forces.

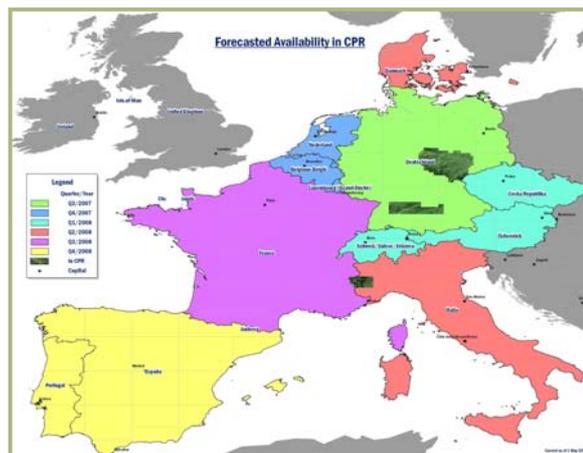


Figure 4: Forecast dates (current, May 2007) for product availability of NEXTMapEurope areas. The various colours starting with light green (September 2007) and ending with yellow (December 2008) show the forecast dates that the data products will be available. Dark green 'thumbnails' are currently in the data base. This forecast may change subject to operational and other priorities.

4. EXAMPLES

In this section we show examples from two potential market areas which should benefit from the availability of NEXTMap coverage. In Figure 5 a flood risk application is exemplified, while in Figure 6 a visualization example is presented. In both cases the DSM is draped by a high resolution colour air photo. Because of the presence of the 1.25 m ORI it is possible to easily ortho-rectify the air photos using rational functions procedures (or similar), making the co-registration of air-photo to DSM relatively simple. The flood risk application involves 3rd-party models for which the DSM product is an important input component. The visualization example relates to many applications and markets ranging from recreation to automotive safety. In this instance it is a scene extracted from a fly-through (Eye-Tour).

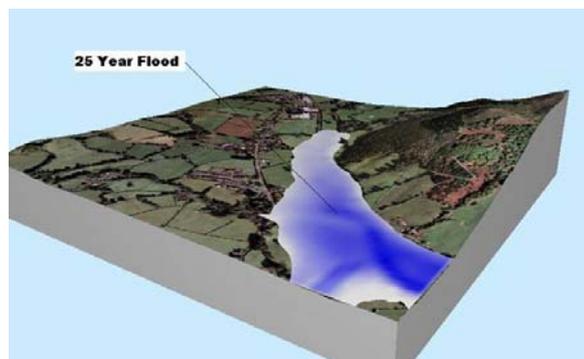


Figure 5: Example from NEXTMap Britain flood risk application

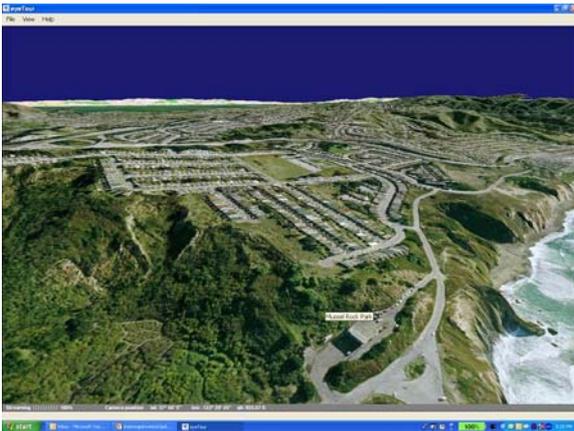


Figure 6: Example from NEXMap USA - California scene extracted from a fly-through near San Francisco

5. THE COMPETITIVE ENVIRONMENT

From a commercial perspective, any technology must evaluate its position in the competitive environment with respect to alternative choices: in short, what does it offer in terms of familiar factors such as price, product specifications, and availability? In particular it is of interest to review the advantages and limitations of DEMs created by space borne sources as well as airborne lidar relative to the NEXMap products described in this paper and relative to markets such as those alluded to in this paper. A brief summary is provided in Table 3. The brown colour suggests areas of comparative disadvantage, at least in the market areas suggested here. Intermap currently does not publish a price list. However, based upon previous published information, it could be within the \$3-\$30/km² regime depending on the factors of quantity, license type and other factors. If this assumption is correct, it appears that it will be complementary rather than competitive with lidar, based upon availability and price. Similarly it appears to be complementary to SRTM based upon the difference in detail provided. The Tandem-X space borne InSAR system will create products of comparable detail to NEXMap if it performs according to specification. However its launch is not scheduled until 2009 and the product is not expected to be available until the 2012-2013 time frame. Therefore it appears that in the context of market, product specification, availability and price the NEXMap products occupy an attractive niche in the competitive environment.

DEM Sources	Sample Spacing (m)	Vertical Accuracy (m) @90%	Availability	Coverage	Price
SRTM	90 30 (USA)	16 abs 10 rel (5.8 ⁽¹⁾)	Now	Global (almost)	'Free'
SPOT 5	30	10 @ abs	Now	Global (almost)	\$ 4/kmsq ⁽²⁾
Tandem-X	10	10 abs 2 rel	2012 -2013	Global	Commercial TBD
NEXMap	5	1.65 abs 1 rel	2008 -2009	USA, W.Europe	Competitive ⁽³⁾ (see text)
Lidar	1 to 3	0.3 ⁽⁴⁾ abs	Now	Local	\$150 - 600/kmsq ⁽⁴⁾

Indicates suitability problems for the markets or applications under consideration

(1) Dean Gesch: comparison with NGS benchmarks in USA
 (2) List price SPOT USA
 (3) Dependant on size, license type, etc
 (4) Nominal - project, location and vendor dependant

Table 3: Comparison of different DEM sources with respect to detail, availability, coverage and price. Note that the vertical accuracy is quoted with respect to the 90% confidence level rather than RMSE in this table.

6. SUMMARY AND CONCLUSIONS

The NEXMap program, based upon airborne InSAR, is in the process of creating a 3-dimensional, homogeneous, seamless database including DSM, DTM and ORI products, for twelve countries of western Europe (in addition to Britain which is already available) and for the USA (excluding Alaska). The DSM is specified at 1 m RMSE vertical accuracy for 5 m sample spacing, while the ORI is specified at 1.25m resolution with better than 2 m RMSE horizontal accuracy. For many applications the combination of detail provided, national and super-national availability and shared pricing through license arrangements, should produce an attractive user opportunity. The creation of the datasets for both areas is well underway and scheduled for completion, in the case of Western Europe, for late 2008 and about 1 year later for the USA. With respect to many market applications such as visualization, flood risk, and auto safety these products should occupy a solid niche, relative to alternative technologies.

REFERENCES

Bamler, R. and P. Hartl, Synthetic aperture radar interferometry, *Inv.Probl.*, vol. 14, pp. R1-R54, 1998

Mercer, B., 2004, DEMs Created from Airborne IFSAR – An Update, *Proceedings of the ISPRS XXth Congress, Istanbul*

Rodriguez, E., and J.M Martin (1992). Theory and design of interferometric synthetic aperture radars. *IEE Proceedings-F, Vol. 139, No. 2, pp. 147-159*

Schwabisch, M. and J. Moreira, 1999. The high resolution airborne interferometric SAR AeS-1. In: *Proceedings of the Fourth International Airborne Remote Sensing Conference and Exhibition/ 21st Canadian Symposium on Remote Sensing, Ottawa, Ontario, Canada.*

Tennant, J. K., and T. Coyne, 1999. STAR-3i interferometric synthetic aperture radar (INSAR): some lessons learned on the road to commercialization. In: *Proceedings of the 4th International Airborne Remote Sensing Conference and Exhibition/21st Canadian Symposium on Remote Sensing, Ottawa, Ontario, Canada, 21-24 June*

Wang, Y., B. Mercer, V.Tao, J. Sharma, and S. Crawford, 2001. Automatic Generation of Bald Earth Elevation Models from Digital Surface Models Created Using Airborne IFSAR., *Proceedings of ASPRS 2001, Washington, DC, CD ROM*

Zebker, Howard A., and J. Villasenor (1992). Decorrelation in Interferometric Radar Echoes., *IEEE Transactions on Geoscience and Remote Sensing, Vol. 30: Number 5, pp 950-959.*

ACKNOWLEDGEMENTS

The author would like to acknowledge his many colleagues at Intermap who have made the NEXMap program a reality.

AUTOMATIC CLASS MEAN CALCULATION OF ROAD SURFACE FROM IKONOS IMAGES USING FUZZY LOGIC AND PARTICLE SWARM OPTIMIZATION

A. Mohammadzadeh^{a,*}, M. J. Valadan Zoej^a, A. Tavakoli^b

^a Geodesy and Geomatics Engineering Faculty, K.N.Toosi University of Technology, No. 1346, Vali_Asr St., Tehran, Iran, Postal Code: 1996715433 - ali_mohammadzadeh@alborz.kntu.ac.ir - valadanzouj@kntu.ac.ir

^b Dept. of Electrical Engineering, Amirkabir University of Technology - tavakoli@aut.ac.ir

KEY WORDS: Fuzzy logic, Particle swarm optimization, IKONOS, Road detection, Class Mean

ABSTRACT:

Automatic road detection from high resolution satellite images has been an active research topic in the past decades. Different solutions are proposed to detect road object such as: fusion-based, fuzzy-based, mathematical morphology, model-based approach, dynamic programming and multi-scale grouping. In this paper, a new fuzzy segmentation method is proposed which is optimized by particle swarm optimization (PSO). The proposed method detects the road network using few samples from its surface. In the IKONOS images, the standard deviation of 10 grey level has been measured for the road classes. In the proposed fuzzy logic system, just one arbitrary pixel up to maximum of three from the road surface is an adequate initial value. The road is identified requiring neither the numbers of the classes nor the corresponding mean values. Particle swarm optimization is used to optimize the proposed fuzzy cost function. The proposed algorithm is applied on real IKONOS satellite image. The results indicate acceptable accuracy for the extracted road surface.

1. INTRODUCTION

Automatic detection and extraction of linear features especially road object has been the interest of many researchers for many years. Diverse solution has been proposed by researchers. Mena (2003) has presented a review of nearly 250 references on automatic road extraction. Furthermore, Zhang (2003), Mena and Malpica (2005), and Quackenbush (2004) have evaluated a wide variety of existing approaches. The applied road extraction methodologies could be summarized as follows (Mohammadzadeh et al., 2006):

- (a) *Road tracking* methods start from a set of seed points and utilize radiometric values of pixels in the tracking process such as profile matching and Kalman filtering (Vosselman and DeKnecht, 1995; Bonnefon et al., 2002; Hu et al., 2004; Kim et al., 2004).
- (b) *Morphological* methods rely on set theory developed by Matheron (1975) and Serra (1982). The method is sensitive to geometry of features and uses set operations such as union, intersection, complementation, dilation, erosion and thinning to identify geometrical characteristics of objects (Zhang et al., 1999; Amini et al., 2002a).
- (c) *Snakes* along with *dynamic programming* is used for extraction of cultural structures such as roads and buildings (Agouris et al., 2001a, b; Péteri et al., 2004). Dynamic programming provides an optimized solution to road model Gruen and Li (1995). Simultaneous use of snakes and dynamic programming is reported by Gruen and Li (1997).
- (d) *Artificial intelligent* approaches do reasoning based on the rules and models like a human being to provide correct, flexible and effective results. Fuzzy logic, neural networks and genetic algorithms are the most popular mathematical tools used in artificial intelligent systems. Fuzzy Logic (Zadeh, 1965) is a multivalued logic that allows intermediate values to be defined between zero and one (Agouris et al., 1998; Amini et al., 2002b). Artificial neural networks simulate the learning process of the

human brain (Doucette et al., 2001; Egmont-Petersena et al., 2002) and genetic algorithms are generally used as an optimisation technique to search the global optimum of a function (Settea and Boullartb, 2001; Brumby et al., 1999).

- (e) *Fusion-based* methods fuse multi-source and multi-type information (Jin and Davis, 2005; Chena et al., 2003; Pajares and de la Cruz, 2004). The information covers satellite images (visible and radar), domain-based models, expert's modeled knowledge, strategies and rules (Pigeon et al., 2001).
- (f) *Multi-scale* along with *multi-resolution analysis* provide an efficient tool for road width detection and analysis of the road across profiles (Mayer et al., 1997; Mayer and Steger, 1998). Multi-resolution analysis is based on the topological relations between roads and other cartographic objects (Hinz et al., 2001; Chen et al., 2004; Hinz and Baumgartner, 2003).

In this research, new image segmentation based on fuzzy logic theory is proposed. Particle swarm optimization (PSO) has been employed to optimize the proposed approach. All of the classification and segmentation methods fell in two important groups: supervised and unsupervised methods. In supervised methods, necessary and enough samples should be selected from most of the possible classes in the image. In this way, classification parameters, like classes' mean and standard deviations are calculated from those selected sample data. The final result of the classification is highly affected by the selected samples and number of the user defined classes. In unsupervised methods, the user must define number of the classes or other equivalent parameters like minimum distance between the classes. But in the proposed method, just one up to three samples from the road surface is sufficient to calculate the mean value of class "road" in the image. Number of the classes and parameters of all the existing classes are not needed. This is an interesting advantage comparing to the previous classification and segmentation methods. By taking different samples from different images, the standard deviation of the road is found to be $\sigma = 10$. It is due to this fact that road

surface is mostly homogenous and high radiometric variations are not expected. To see the efficiency of the method, it is applied on both simulated and real pansharpened IKONOS satellite image.

2. FUZZY SEGMENTATION OPTIMIZED BY PSO

Mohammadzadeh et al. (2006) proposed a new fuzzy segmentation method, which computes road radiometric mean values in each band using just few samples from its surface. In the proposed system, a blind search with one gray level accuracy in the search domain is performed to find the best mean value. The search domain is considered to be discrete and therefore few point are tested and there is fear of incorrect answers. Due to the mentioned reasons, a global minimum search approach should be applied to the proposed method. In addition, this will lead to a higher accuracy with sub gray level accuracy. In other words, PSO will find the global minimum of the fuzzy cost function with higher accuracy. Figure 1 illustrates the flowchart of the proposed algorithm.

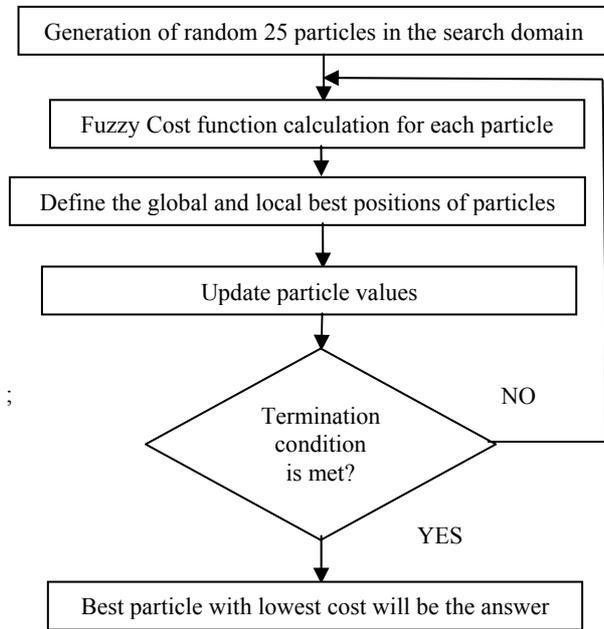


Figure 1. Flowchart of the method

The implemented PSO generates $m=25$ random particles in a continuous cubic search domain. Each of the particles has three values in the three dimensional search space, which are possible mean values of road in bands 1, 2, and 3. After taking few samples from the road surface, maximum and minimum values of particle 'm' in the band 'n' ($p_{m,n}^{old}$) will be defined as :

M : user selected sample pixel;
 $M = [M_1, M_2, \dots, M_n, \dots, M_N]$: A pixel from the image
 $\forall M_n \in X, p_{m,n}^{old} = [(M_n - 15), (M_n + 15)] ; n = 1, 2, 3$

Then the fuzzy cost function is evaluated for each of the produced particles. The particle that has the minimum cost value will be chosen as the best particle in each iteration. Then position of the particles will be updated based by the velocity vector which are influenced by both the best global solution associated with the lowest cost ever found by a particle and the best local solution associated with the lowest cost in the present population (see formulas 1 and 2). The best local solution

replaces the best global solution if it has lower cost. The above procedure is carried out until the termination condition is met. For example, if differences of the two consequently calculated best mean values were below a user-defined threshold then the process will be finished and the last global best position would be considered as the final calculated best mean values of the road.

$$v_{m,n}^{new} = v_{m,n}^{old} + \Gamma_1 \times r_1 \times (p_{m,n}^{local\ best} - p_{m,n}^{old}) + \Gamma_2 \times r_2 \times (p_{m,n}^{global\ best} - p_{m,n}^{old}) \tag{1}$$

$$p_{m,n}^{new} = p_{m,n}^{old} + v_{m,n}^{new} \tag{2}$$

Where,

$v_{m,n}^{old}$ = particle old velocity

$v_{m,n}^{new}$ = particle updated velocity

$p_{m,n}^{old}$ = particle old position

$p_{m,n}^{new}$ = particle updated position

r_1, r_2 = independent uniform random numbers

$\Gamma_1 = \Gamma_2 = 2$ learning factor

$p_{m,n}^{local\ best}$ = best local solution

$p_{m,n}^{global\ best}$ = best global solution

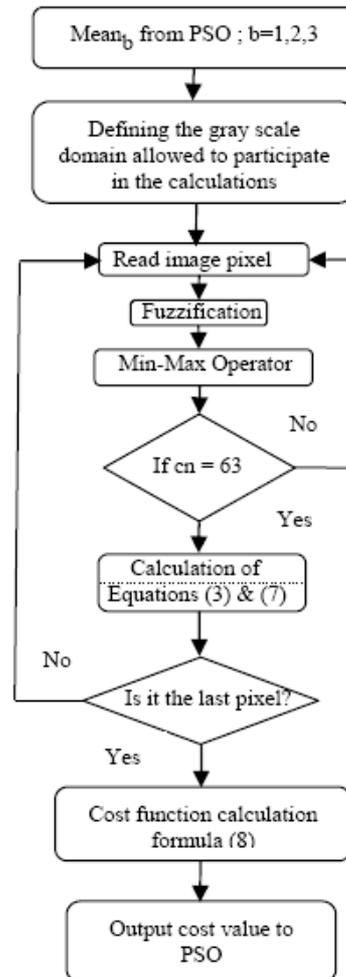


Figure 2. Flowchart of the fuzzy method

Figure 2 shows the flowchart of the implemented fuzzy system. The particle position as $p_{m,n}^{old}$ from the PSO method represents $mean_b$; $b=1,2,3$ which are the inputs of the fuzzy system. The pixels should satisfy the following criteria to participate in the fuzzy process:

$$X=[x_1, x_2, \dots, x_b, \dots, x_B]; \text{ A pixel from the image} \quad (3)$$

$$\forall x_b \in X, (mean_b - 5.25 \cdot \sigma) \leq x_b \leq (mean_b + 5.25 \cdot \sigma) ; b=1,2,3$$

There are $5 \cdot 5 \cdot 5 = 125$ possible fuzzy classes for three bands. Figure depicts the generated $5 \cdot 5 = 25$ classes in two bands.

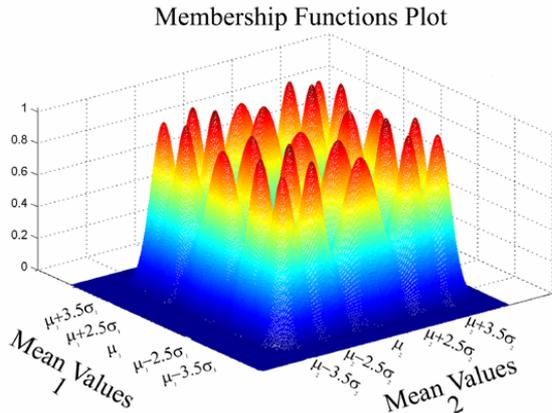


Figure 3. Defined classes using MFs of bands 1 and 2

Matrix F is calculated for each pixel of the image (see equation 4). The maximum among the minimum value of each column is found and its column number is attributed "cn". If $cn \neq 63$, then the image pixel is considered as non-road.

$$f_{b,c}(x_b) = \exp\left(-\frac{(x_b - \mu_{b,c})^2}{2 \cdot \sigma_{b,c}^2}\right) \Rightarrow f = [f_{b,c}] \quad (4)$$

$$\forall x_b \in [0, 255] \sum_{i=1}^{c=125} f_{b,i}(x_b) = 1$$

If "t" is the total number of initially identified road pixels, then for a pixel X_j where $1 \leq j \leq t$:

$$r_j = \begin{bmatrix} r_{1,1} \\ r_{2,1} \\ r_{3,1} \end{bmatrix}_{(M_j)} = \begin{bmatrix} (x_1 - \mu_{1,cn}) \cdot \left(\frac{1 - f_{1,cn}}{\text{sum}(f_{1,1:125})}\right) \\ (x_2 - \mu_{2,cn}) \cdot \left(\frac{1 - f_{2,cn}}{\text{sum}(f_{2,1:125})}\right) \\ (x_3 - \mu_{3,cn}) \cdot \left(\frac{1 - f_{3,cn}}{\text{sum}(f_{3,1:125})}\right) \end{bmatrix}_{(M_j)} \quad (5)$$

$$\text{sig_cn} = \begin{bmatrix} \sigma_{1,cn}^2 & 0 & 0 \\ 0 & \sigma_{2,cn}^2 & 0 \\ 0 & 0 & \sigma_{3,cn}^2 \end{bmatrix}$$

$$c_j = \sqrt{r_j \cdot \text{inv}(\text{sig_cn}) \cdot r_j^T} \Rightarrow \text{sum_c} = \sum_{j=1}^t c_j \quad (6)$$

$$d_j = \sum_{b=1}^3 \left(1 - \frac{f_{b,cn}}{\text{sum}(f_{b,1:125})}\right)^2 \Rightarrow \text{sum_d} = \sum_{j=1}^t d_j \quad (7)$$

This procedure is applied to all image pixels (or a selected window) and the cost function is calculated as:

$$\text{Cost_mean} = \frac{\text{sum_c}}{\text{sum_d} \cdot t} \quad (8)$$

The value of 'cost_mean' will be sent back to the PSO algorithm as the cost value for a particle like $p_{m,n}^{old}$.

3. IMPLEMENTATION AND RESULTS

Four simulated patterns are generated with hypothetic mean and standard deviation values as defined in Table 1 and is shown Figure 4. Table 2 shows the computed mean values for the simulated patterns using the proposed fuzzy-PSO method. The sub gray level difference between the computed mean values and the corresponding hypothetic ones indicate the efficiency and accuracy of the proposed method.

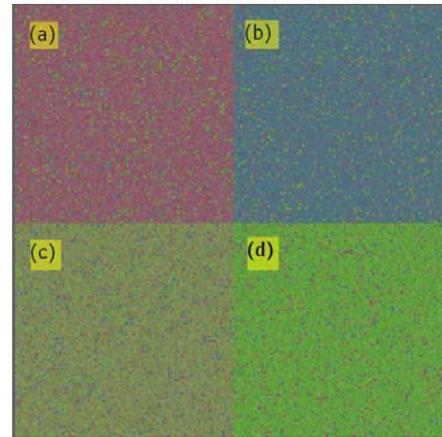


Figure 4. Simulated patterns (a) to (d).

Table 1. Hypothetic values used for pattern generation.

Pattern	Band 1 (grey level)		Band2 (grey level)		Band3 (grey level)	
	μ_1	σ_1	μ_2	σ_2	μ_3	σ_3
Pattern (a)	140.3	6	93.2	9	104.8	8
Pattern (b)	82.7	7	113.7	6	125.9	4
Pattern (c)	123.3	8	137.5	10	86.4	7
Pattern (d)	98.7	10	158.3	9	63.1	6

Table 2. Computed mean values.

Pattern	Band 1 (grey level)	Band2 (grey level)	Band3 (grey level)
	μ_1	μ_2	μ_3
Pattern (a)	139.90	93.23	104.82
Pattern (b)	82.98	113.29	125.88
Pattern (c)	123.79	137.87	86.18
Pattern (d)	98.98	157.95	62.94

Figure 5 shows a window of $595 \cdot 425$ pixels from a pan-sharpened IKONOS image of Kish Island in Iran. In large scenes the road surface would be itself consist of two or three

subclasses. Two subclasses were identified using the proposed methodology. The first arbitrary input sample was (63, 81, 77) and the calculated road mean for the first subclass was (68.30, 90.45, 85.06). The second arbitrary input sample was (100, 120, 100) and the calculated road mean for the first subclass was (93.66, 109.72, 95.79). The result of applying the proposed method is shown in Fig. 6.



Figure 5. IKONOS pan-sharpened image

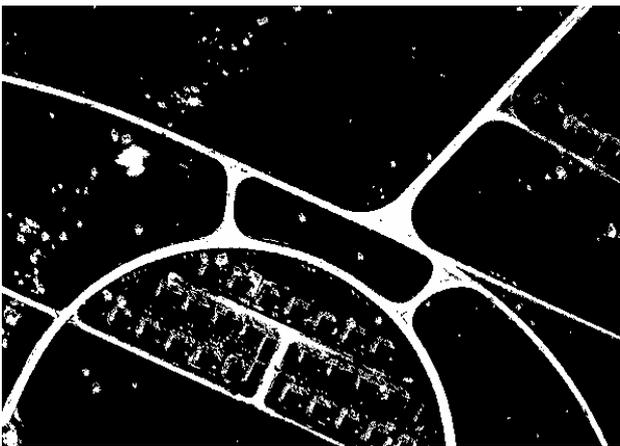


Figure 6. Detected road surface using fuzzy-PSO method.

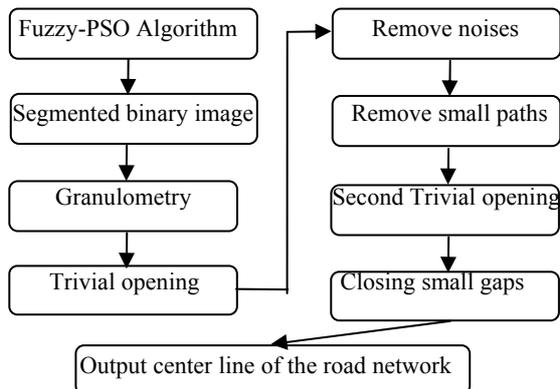


Figure 7. Block diagram used in morphology approach

Mohammadzadeh et al. (2004) extract road centreline based on mathematical morphology. Figure 7 shows the block diagram used in mathematical morphology. Mathematical morphology is non-linear image processing tool, which is sensitive to the objects shape. As roads appear as elongated narrow object in

the image, therefore it is suitable to use mathematical morphology to extract road centreline. After applying the mentioned morphological algorithm, the final extracted road is shown in figure 8.



(a) result of trivial opening



(b) Closing operation



(c) result of opening



(d) applying second trivial opening



(e) second closing operator



(f) extracted road centerline

Figure 8. Road extraction from pan-sharpened IKONOS image of a rural area of Kish island: (a) result of trivial opening, (b) closing operator, (c) result of opening operation, (d) applying second trivial opening, (e) second closing operator, and (f) extracted road centerline in red which is superimposed on the original image.

4. CONCLUSIONS

The proposed fuzzy-PSO approach calculates the road's mean values automatically in different bands by using one up to three samples from the road surface with sub grey level accuracy. It

is tested on both simulated and high resolution satellite images. Mathematical morphology is used to extract the road centerline. We should consider that the standard deviation of the road class is considered $\sigma = 10$ according to different sample we had from different IKONOS images. It is due to this fact that road appear as homogenous surface in the satellite images and so we do not expect high variance in the road surface reflectance data. The proposed algorithm shows a promising result. But one of our future efforts is to find a solution to obtain the exact standard deviation of the road surface from the satellite image in an automatic manner to achieve a more reliable detection algorithm. This would also lead us to detect other planimetric objects.

5. REFERENCE

- Agouris, P., Gyftakis, S. and Stefanidis, A., 1998. Using a fuzzy supervisor for object extraction within an integrated geospatial environment. *International Archives of Photogrammetry and Remote Sensing*, 32(3/1):191-195.
- Agouris, P., Gyftakis, S. and Stefanidis, A., 2001a. Quality-aware deformable models for change detection. *Proceeding of IEEE International Conference on Image Processing (ICIP)*, Thessaloniki, Greece. 2: 805-808.
- Agouris, P., Stefanidis, A. and Gyftakis, S., 2001b. Differential Snakes for Change Detection in Road Segments. *Photogrammetric Engineering & Remote Sensing*, 67(12): 1391-1399.
- Amini, J., Saradjian, M.R., Blais, J.A.R., Lucas, C. and Azizi, A., 2002a. Automatic road side extraction from large scale image maps. *International Journal of Applied Earth Observation Geoinformation*, 4(2): 95-107.
- Amini, J., Lucas, C., Saradjian, M.R., Azizi, A. and Sadeghian, S., 2002b. Fuzzy logic system for road identification using Ikonos images. *Photogrammetric Record*, 17(99): 493-503.
- Bonnefon, R., Dherete, P. and Desachy, J., 2002. Geographic information system updating using remote sensing images. *Pattern Recognition Letter*, 23 (9): 1073-1083.
- Brumby, S. P., Perkins, S. J., Theiler, J., Szymanski, J. J., Bloch, J. J., and Mitchell, M., 1999. Investigation of image feature extraction by a genetic algorithm. *Proceedings of the International Society for Optical Engineering, SPIE 3812*, Bellingham. Pages: 24-31.
- Chen, T., Wang, J., Zhang, K., 2004. A Wavelet Transform Based Method for Road Centerline Extraction. *Photogrammetric Engineering & Remote Sensing*, 70(12): 1423-1432.
- Chena, C.-M., Hepnerb, G.F. and Forster, R.R., 2003. Fusion of hyperspectral and radar data using the HIS transformation to enhance urban surface features. *ISPRS Journal of Photogrammetry & Remote Sensing*, 58(1-2): 19-30.
- Doucette, P., Agouris, P., Stefanidis, A. and Musavi, M., 2001. Self-Organized Clustering for Road Extraction in Classified Imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 55(5-6): 347-358.

- Egmont-Petersena, M., De Ridderb, D. and Handelse, H., 2002. Image processing with neural networks-a review. *Pattern Recognition*, 35(10): 2279–2301.
- Gruen, A. and Li, H., 1995. Semi-automatic road extraction by dynamic programming. *ISPRS Journal of Photogrammetry and Remote Sensing*, 50(4):11-20.
- Gruen, A. and Li, H., 1997. Semi-automatic linear feature extraction by dynamic programming and LSB-Snakes. *Photogrammetric Engineering & Remote Sensing*, 63(8): 985-995.
- Hinz, S. and Baumgartner, A., 2003. Automatic extraction of urban road networks from multi-view aerial imagery. *ISPRS journal of photogrammetry & remote sensing*, 58(1-2): 83-98.
- Hinz, S., Baumgartner, A., Mayer, H., Wiedemann, C., Ebner, H., 2001. Road extraction focusing on urban areas. *Automatic Extraction of Man-Made Objects from Aerial and Space Images (III)* (Eds. Baltsavias, E., Gruen, A. and Van Gool, L.). Balkema Publishers, Rotterdam, 415 pages: 255-265.
- Hu, X., Zhang, Z. and Tao, C.V., 2004. A Robust Method for Semi-Automatic Extraction of Road Centerlines Using a Piecewise Parabolic Model and Least Square Template Matching. *Photogrammetric engineering and remote sensing*, 70(12): 1393-1398.
- 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(4): 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): 1417-1422.
- Matheron, G., 1975. Random set and integral geometry. John Wiley & Sons Inc., New York. 288 pages.
- Mayer, H., Laptev, I., Baumgartner, A. and Steger, C., 1997. Automatic road extraction based on multi-scale modeling, context and snakes. *International Archives of Photogrammetry and Remote Sensing*, 32(3-2W3):106-113.
- Mayer, H. and Steger, C., 1998. Scale-space events and their link to abstraction for road extraction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 53 (2): 62– 75.
- Mena, J.B., 2003. State of the art on automatic road extraction for GIS update: a novel classification. *Pattern recognition letters*, 24(16): 3037-3057.
- Mena, J.B. and Malpica, J.A., 2005. An automatic method for road extraction in rural and semi-urban areas starting from high resolution satellite imagery. *Pattern Recognition Letters*, 26(9): 1201–1220.
- Mohammadzadeh, A., Tavakoli, A., Valadan Zoej, M.J., 2004. Automatic linear feature extraction of Iranian roads from high resolution multi-spectral satellite imagery. *International Archives of Photogrammetry and Remote Sensing*, Istanbul, Turkey, PS ThS 12: Automated Object Extraction and Computer Vision Application, July, 13:30 - 15:30.
- Mohammadzadeh, A., Tavakoli, A., Valadan Zoej, M.J., 2006, “Road extraction based on fuzzy logic and mathematical morphology from pan-sharpened IKONOS images”, The photogrammetric record, Vol. 21 (113):44-60.
- Pajares, G., De la Cruz, J.M., 2004. A wavelet-based image fusion tutorial. *Pattern Recognition*, 37(9): 1855– 1872.
- Péteri, R., Couloigner, I. and Ranchin, T., 2004. Quantitatively Assessing Roads Extracted from High-Resolution Imagery. *Photogrammetric Engineering & Remote Sensing*, 70(12): 1449–1456.
- Pigeon, L., Solaiman, B., Toutin, T. and Thomson, K., 2001. Linear planimetric feature domains modeling for multisensor fusion in remote sensing. *Proceedings of SPIE Aero Sense 2000. SPIE*, 4051. 8 pages (CD-ROM).
- Quackenbush, L. J., 2004. Review of Techniques for Extracting Linear Features from Imagery. *Photogrammetric engineering and remote sensing*, 70(12): 1383-1392.
- Serra, J., 1982. *Image analysis and mathematical morphology*. Academic Press, London. 610 pages.
- Serra, J. and Vincent, L., 1992. An overview of morphological filtering. *Circuits, Systems and Signal Processing*. 11(1):47-108.
- Settea, S. and Boullart, L., 2001. Genetic programming: principles and applications. *Engineering Applications of Artificial Intelligence*, 14(6): 727–736.
- Vosselman, G. and De Knecht, J., 1995. Road tracing by profile matching and Kalman filtering. *Automatic Extraction of Man-Made Objects from Aerial and Space Images* (Editors: Gruen, A., Kuebler, O. and Agouris, P.). Birkhäuser, Basel, Switzerland. 321 pages: 265-274.
- Zadeh, L.A., 1965. *Fuzzy sets. Inform. Control*, 8: 338-353.
- Zhang, C., 2003. Updating of cartographic road database by image analysis. Ph.D. thesis, Institute of Geodesy and Photogrammetry, ETH Zurich. 188 pages.
- Zhang, C., Murai, S. and Baltsavias, E. P., 1999. Road network detection by mathematical morphology. *Proceedings of ISPRS Workshop on 3D Geospatial Data Production: Meeting Application Requirements*, Paris, France. Pages: 185-200.

MOVING TARGETS VELOCITY AND DIRECTION ESTIMATION BY USING A SINGLE OPTICAL VHR SATELLITE IMAGERY

M. Pesaresi^{a,*}, K. Gutjahr^b, E. Pagot^a

^a European Commission, Joint Research Centre, Institute for the Protection and Security of the Citizen (IPSC), via E.Fermi 1 TP 267 Ispra 21020 (VA), Italy- Martino.Pesaresi@jrc.it

^b Joanneum Research ForschungsgesmbH, Institute of Digital Image Processing, Wastiangasse 6, A - 8010 Graz, Austria

Commission VI, WG VI/4

KEY WORDS: Fast moving target detection, satellite imagery, optical very high resolution (VHR) sensors

ABSTRACT:

The present contribution demonstrates the feasibility and explores the limits of a new method for estimating the velocity and direction of moving targets using a single VHR satellite dataset. The method is based on the fact that there is a time lag between the data collection of the panchromatic (PAN) and multi-spectral (MS) sensors in the same VHR platform. Consequently, it is developed around three main steps: i) accurate image-to-image registration between MS and PAN images with a sub-pixel displacement error, ii) precise location of barycentre of targets by mathematical morphology-based image transforms, and iii) estimation of the targets ground velocity and direction using the MS-PAN spatial displacement, the known time lag, and an image-to-ground transformation taking into account the interior and exterior orientation of the sensors and a terrain height reference. An evaluation of the reliability and limits of the proposed method based on the observation of the results regarding manually-selected moving and non-moving targets is included.

1. INTRODUCTION

Most of the applications using VHR data assume that the collection of one imagery data set occurs at one point in the time line. If a multi-temporal analysis is needed, this requires a multiple collection of imagery representing the same spatial domain at different points of the time line. Then change detection techniques are usually applied in order to enhance evolving phenomena or detect moving targets. The associated intrinsic limit is to be able to observe only changes having a frequency of occurrence/change or a velocity on the ground much slower than the revisit time of the satellite platforms. These are of the order of some days in optimal conditions of cloud coverage and other tasks constraints.

The detection of moving targets with remotely-sensed data is

usually done using conventional radar or imaging radar techniques, such as SAR (Meyer et al., 2006; Pettersson, 2004) and along-track SAR interferometry approaches have been proposed for traffic monitoring using SRTM/X-SAR ATI data (Suchandt et al., 2006). In alternative, optical video cameras are used for traffic monitoring (Munno et al., 1993), but they are mostly based on ground or aircraft platforms that offer technical characteristics as image spatial resolution and frequency of image collection that unfortunately are not available in the satellite platforms at the date (Toth and Grejner-Brzezinska, 2006; Reinartz et al. 2006). For security applications aircraft platforms have also limitations related to the accessibility of remote areas especially relevant in case of conflict-prone places or areas with severe security-related concerns that are not accessible by civilian missions.

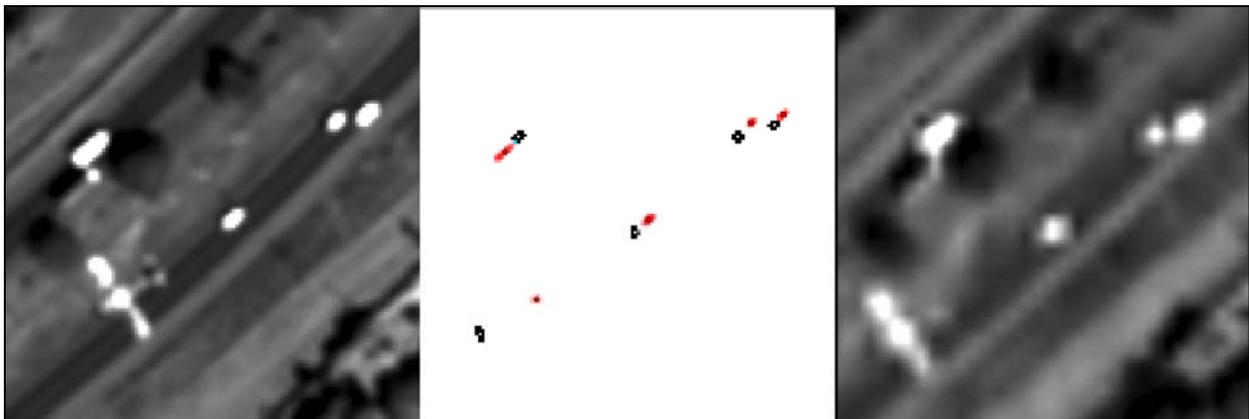


Figure 1 – sub-sample of the PAN (left) and MS (right) data used for detection of the image blobs and precise barycentres (center). In red and black the blobs coming from the PAN and MS sensors, respectively.

* Corresponding author.

The proposed methodology changes radically the perspective by exploiting the fact that in reality satellite VHR optical systems have two different sensors on board, respectively multi-spectral (MS) and panchromatic (PAN), which do not capture the information instantaneously but with a given time lag between the two data collections. By knowing the exact time lag between the two data collections, it is possible to estimate the target's ground velocity with the approximation of the pixel size. The shift of fast-moving features in the PAN and MS images produced by the time lag between the two image data collections are well known characteristics of VHR data (Dial 2003), but no systematic evaluation of the reliability of this information for estimation of the target velocity is available in the literature at the date. Usually this characteristic of VHR satellite data is considered as an artefact to be fixed for improving the image-to-image matching and digital surface extraction from stereo-pairs (Baltsavias and others 2001, Jacobsen 2005).

The detection of moving targets using remotely-sensed data has a clear connection with the support to all the activities related to traffic monitoring, estimation of traffic volume, and the modelling and planning of the traffic both in the terrestrial and maritime environment. An important application field can also be related to security and defence, where it is important to know if, and possibly where, there are moving vehicles or in general targets in the scene under analysis, with the estimation of ground velocity and direction. Moving vehicles detection using remotely sensed data could be also important in case of necessity to estimate the illegal border crossing of goods and people on remote areas difficult to control with traditional approaches.

2. THE PROPOSED METHOD

2.1 Processing flow

The concept presented here is tested using a part of VHR data coming from the Quick Bird® satellite of Digital Globe, having nominally 0.6 and 2.4 meter spatial resolution of the panchromatic and multi-spectral sensor, respectively. The processing flow consists on six steps as follow:

1. Fusion of the MS channels in order to simulate the spectral response of the panchromatic P
2. MS and P resolution matching
3. Image-to-image registration by correlation matching of the MS output of point (2) on the panchromatic channel using the latter as reference
4. Rough selection of potentially interesting target points (fuzzy markers)
5. Automatic refinement of the targets location
6. Estimation of the targets ground velocity

Steps 1 and 2 are functional to the preparation of the image-to-image registration, with the objective to use the multi-spectral data (MS) of the sensor for creating an MS-derived image as similar as possible to the spectral characteristics of the panchromatic (P) image. For Quickbird images the fusion is done by spectral averaging all the values of the MS sensor, simulating the frequency response of the P sensor. To ensure same gray value distributions a histogram matching is applied to the fused MS-image with the panchromatic image serving as reference. In order to reach the resolution of the P sensor, the MS-derived image was then over-sampled by a factor 4 through

bi-linear resampling approach. The original P image was filtered by a 4 x 4 low-band-pass convolution filter in order to create an image with similar spatial characteristics to the MS-derived image, but without losing the accuracy on the location of the barycentre of the targets.

Steps 3, 4, 5 are described in the following paragraph 4, and the final step 6 is detailed in paragraph 5.

2.2 Accuracy considerations

The time lag $\Delta t_{Pan,MS}$ between the MS and P sensor acquisition is the limiting factor for the velocity determination. Obviously a coarse estimation can be found according to

$$v = \frac{\Delta d_{Pan,MS}}{\Delta t_{Pan,MS}} \quad \text{where } \Delta d_{Pan,MS} \text{ is the length of the}$$

displacement vector between the target locations in the MS and P image. In case of Quickbird, the time lag is about 0.2 sec and the nominal spatial resolution of the panchromatic and multi-spectral sensor is 0.6 and 2.4 meters, respectively. Thus a target velocity of 50 km/h corresponds to 4.7 pixels in the P image or, the other way around, a displacement of 1 pixel corresponds to about 10 km/h.

As a consequence reliable velocity results can only be achieved if sub-pixel accuracy is ensured throughout the processing flow, most particularly for the image-to-image registration and the target localization. Since the velocity can be estimated from relative displacements, the absolute location accuracy and the quality of the reference height information are of minor importance.

3. MOVING TARGET DETECTION

As the accuracy of the estimated velocity depends on the quality of the measured displacement vectors, an image-to-image registration is required. Therefore, a dense grid of points in the modified P image is identified with the fused MS image, applying image matching techniques (Paar and Pölzleitner, 1992). These points are used to calculate an affine transformation between both images. Poor identification results are filtered by the back-matching distance and additionally can be removed during the polynomial calculation by looking at the individual point residuals.

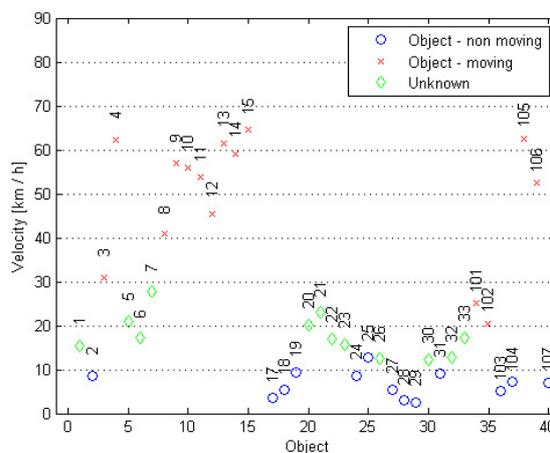


Figure 2 – estimated velocity for the selected targets (objects)

We chose a grid of 10 by 10 pixels in the area of interest (1000 by 1000 pan-pixels) yielding about 9000 tie-points (TP's) candidates. These points were first filtered using a threshold of 0.5 pixels for the back-matching distance. When fitting an affine polynomial to the remaining 1901 points again 198 tie-points were eliminated because of their poor residual statistic. The residual statistics of the polynomial fitting is summarized in Table 1. Finally, the fused MS image is warped to the P image using the found affine transformation.

1703 TP's	Column [pxl]	Line [pxl]	Length [pxl]
<i>Std.Dev.</i>	0.76	0.81	1.11
<i>Mean</i>	0.00	0.00	0.99
<i>Minimum</i>	-1.89	-2.02	0.04
<i>Maximum</i>	1.87	2.02	2.59

Table 1: Residual statistics of the polynomial transformation calculation

The rough selection of potentially interesting target points on step 4 of the processing flow has been solved by visual inspection of the images after the image-to-image automatic matching. The procedure was the following: a disk of radius r overlapping interesting targets (moving and stable objects) was manually digitised and recorded in a separate image layer. Stable targets were acquired in order to have an indirect measure of the image-to-image registration intrinsic error. The disk serves as fuzzy marker with a spatial fuzziness proportional to parameter r . In this case $r=15$ pixels (corresponding to a radius of 9 meters on the ground) was chosen in order to include any potential interesting target such as cars and/or trucks.

The precise location of the targets (step 5) was calculated using a procedure based on mathematical morphology and involving the detection of the image structures (blobs) related to the manually-fuzzy-selected targets together with the calculation of the precise barycentre of these image blobs.

In order to detect both targets brighter and darker than the

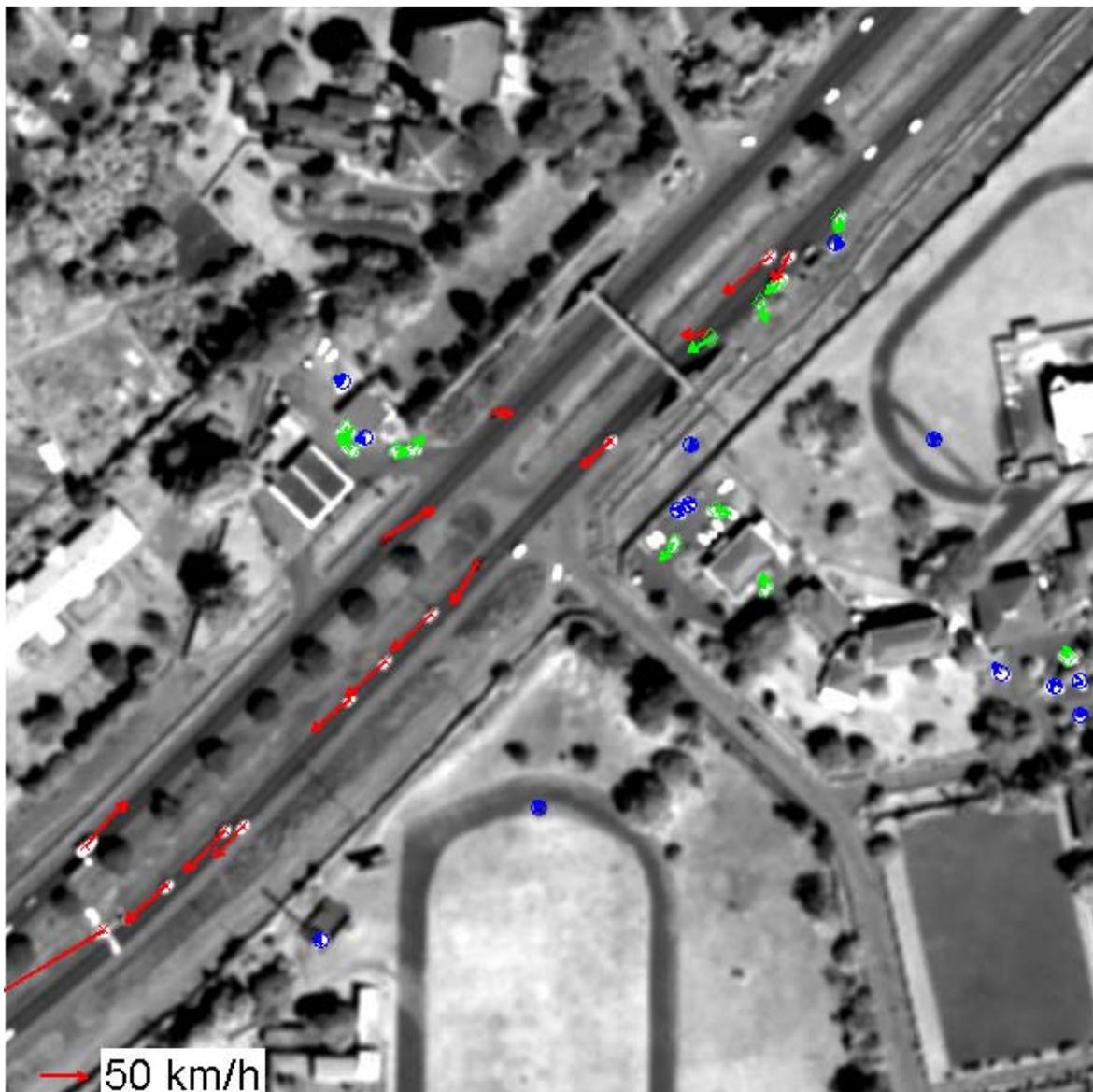


Figure 3 – estimated velocity of targets overlaid on the PAN image. Red, green and blue targets were classified as “moving”, “unknown”, and “non-moving”, respectively, by visual inspection.

surrounding image background, the procedure used a parallel approach based on the calculation of the residuals of morphological opening and morphological closing by reconstruction, respectively.

In particular, the image blobs of brighter targets of the image $T_{bright}(f)$ are calculated by the white top-hat by reconstruction (RWTH) morphological transform as follows:

$$\begin{aligned} T_{bright}(f) &= RWTH(f) > 0 \\ f - \gamma_R^{(n)}(f) &> 0 \end{aligned} \quad (1)$$

While the image blobs corresponding to the darker targets of the image $T_{dark}(f)$ are calculated by the black top-hat by reconstruction (RBTH) morphological transform as follows:

$$\begin{aligned} T_{dark}(f) &= RBTH(f) > 0 \\ \phi_R^{(n)}(f) - f &> 0 \end{aligned} \quad (2)$$

Where $\gamma_R^{(n)}(f)$ and $\phi_R^{(n)}(f)$ are the opening and closing by reconstruction, respectively, with a structuring element of size (n) of the input image (f) (Soille 2003, p.212). The target blobs were derived from the support area of the $RWTH(f)$ and $RBTH(f)$, defined as the region of pixels in the transformed image $T_{bright}(f)$ or $T_{dark}(f)$, having residuals greater than zero. The size (n) of the structuring element used here is of course linked to the size of the targets we want to detect as bright or dark structure in the image: in this case the minimal 3x3 box was sufficient.

The barycentre of the detected blobs is then calculated by simply averaging the x and y image coordinates of all pixels belonging to each specific blob. Other alternatives have been tested as the detection of the extrema (regional maxima and regional minima for bright and dark blobs, respectively) of the grey-level functions associated with the blobs, or the weighted average of the coordinates by the intensity of the blob contrast. However, by analysing the estimated displacement of stable targets, the results of these more complicated procedures were not reducing this displacement but increasing it by a factor of about 5 to 10 % with respect to what we obtained with the simple coordinate averaging method. Therefore, the last method was chosen in the following.

4. ESTIMATION OF THE TARGETS' GROUND VELOCITY

4.1 Ground-to-image and inverse transformation

To determine the velocity (step 6 of the processing flow), the barycentre pixel coordinates of the detected blobs in the pre-processed MS and PAN images had to be geo-referenced. This was based on a so-called image-to-ground transformation which requires the additional information on the interior and exterior orientation of the sensors at the acquisition time of the image and a height reference. In case of Quickbird or Ikonos sensors, the rational polynomial coefficients (RPC's) are supplied with the image data and directly describe the ground-to-image transformation by two fractions of cubic polynomials:

$$x = \frac{f_x(X, Y, Z)}{g_x(X, Y, Z)} \quad \text{and} \quad y = \frac{f_y(X, Y, Z)}{g_y(X, Y, Z)} \quad (3)$$

The original sensor polynomials depend on geographic coordinates (λ, φ, h) but we prefer the representation in an Earth-centred, Earth-fixed Cartesian system (X, Y, Z).

4.1.1 Image-to-map transformation

Calculation of East and North ground coordinate for given height. Approximations for East and North:

$$\begin{aligned} \bar{r} &= (r - r_0) / \mu_r \\ \bar{r} &= \frac{L_{1,1} + L_{1,2} \cdot \bar{\lambda} + L_{1,3} \cdot \bar{\phi} + L_{1,4} \cdot \bar{H}}{L_{2,1} + L_{2,2} \cdot \bar{\lambda} + L_{2,3} \cdot \bar{\phi} + L_{2,4} \cdot \bar{H}} \end{aligned} \quad (4)$$

$$\begin{aligned} \bar{c} &= (c - c_0) / \mu_c \\ \bar{c} &= \frac{C_{1,1} + C_{1,2} \cdot \bar{\lambda} + C_{1,3} \cdot \bar{\phi} + C_{1,4} \cdot \bar{H}}{C_{2,1} + C_{2,2} \cdot \bar{\lambda} + C_{2,3} \cdot \bar{\phi} + C_{2,4} \cdot \bar{H}} \end{aligned} \quad (5)$$

Solve for $\bar{\phi}$ and $\bar{\lambda}$:

$$\begin{aligned} (L_{1,2} - \bar{r} \cdot L_{2,2}) \cdot \bar{\lambda} + (L_{1,3} - \bar{r} \cdot L_{2,3}) \cdot \bar{\phi} &= \\ (L_{2,1} + L_{2,4} \cdot \bar{H}) \cdot \bar{r} - (L_{1,1} + L_{1,4} \cdot \bar{H}) \end{aligned} \quad (6)$$

$$\begin{aligned} (C_{1,2} - \bar{c} \cdot C_{2,2}) \cdot \bar{\lambda} + (C_{1,3} - \bar{c} \cdot C_{2,3}) \cdot \bar{\phi} &= \\ (C_{2,1} + C_{2,4} \cdot \bar{H}) \cdot \bar{c} - (C_{1,1} + C_{1,4} \cdot \bar{H}) \end{aligned} \quad (7)$$

4.1.2 Inverse normalization to get ϕ and λ .

Iterate for East (X) and North (Y):

$$\begin{aligned} r &= r(\bar{X} + d\bar{X}) \\ r &= r(\bar{X}_0) + \frac{\partial r}{\partial \lambda} \cdot d\lambda + \frac{\partial r}{\partial \phi} \cdot d\phi + \frac{\partial r}{\partial H} \cdot dH \end{aligned} \quad (8)$$

$$\begin{aligned} c &= c(\bar{X} + d\bar{X}) \\ c &= c(\bar{X}_0) + \frac{\partial c}{\partial \lambda} \cdot d\lambda + \frac{\partial c}{\partial \phi} \cdot d\phi + \frac{\partial c}{\partial H} \cdot dH \end{aligned} \quad (9)$$

$$\frac{\partial r}{\partial \lambda} \cdot d\lambda + \frac{\partial r}{\partial \phi} \cdot d\phi + \frac{\partial r}{\partial H} \cdot dH = r - r(\bar{X}_0) \quad (10)$$

$$\frac{\partial c}{\partial \lambda} \cdot d\lambda + \frac{\partial c}{\partial \phi} \cdot d\phi + \frac{\partial c}{\partial H} \cdot dH = c - c(\bar{X}_0) \quad (11)$$

4.1.3 Calculus of the solutions

The solution of the inverse transformation can be found iteratively using Taylor series expansions. It is a well known and documented problem that the meta-information has a limited and variable absolute accuracy (Jacobsen 2005, Dial and Grodecki, 2002). Generally a sensor model optimization based on ground control points (GCP's) is required. This is crucial for security related applications as the access to the area of interests is restricted in many cases. Anyhow, the accuracy of the image-to-ground transformation also depends on the used

reference height. We used the globally available SRTM C-band DEM with a nominal resolution of about 90 m and a relative height accuracy of 10 m. Obviously this level of detail and height accuracy does not meet the requirements specified above.

In our approach we try to solve, or at least release, all these problems in the image-to-image registration step. First this registration reduces the sensor geometries to be considered and optimized to one (pan-chromatic sensor). Secondly we again have to stress that we are only interested in relative displacements. Thus the absolute pointing accuracy is of minor interest if the topography around the moving objects is changing smoothly. This is a feasible condition for vehicles as the road networks in general do not show very steep height gradients.

4.1.4 Results

Figure 2 shows the velocity estimations of 40 targets that were classified by visual inspection. Three classes were assigned: "non moving objects" such as for parked vehicles or trees, "moving objects", and "unknown" when it was hard for the interpreter to decide whether or not the target was moving. It is evident that the "non-moving" targets are almost all clustered below the velocity line of 10km/h, which corresponds to roughly one panchromatic pixel of displacement with the given MS-PAN collection time lag (see point 3.2). This can be considered as the technological limit of the tested source given by the constraints on available resolution, MS and PAN image matching precision, and MS-PAN collection time lag. The targets with estimated velocity between 10 km/h and 30 km/h were almost "unknown" from the visual inspection, because the displacement was too small to be detected univocally by manual means, while the targets with estimated velocity greater than 30 Km/h were all labelled as "moving targets" by the visual inspection.

Moreover, if we check the direction and angle together with the velocity of the moving targets on the road, we can observe that the results are coherent with the expected opposite directions in the two road lanes (Figure 3).

5. CONCLUSIONS

The present contribution demonstrates the feasibility and explores the limits of a new method for estimating the velocity and direction of moving targets using a single VHR satellite dataset. The method is based on the fact that there is a time lag between the data collection of the PAN and MS sensor in the same VHR platform, and it is developed around three main steps: i) image-to-image registration between MS and PAN images, ii) precise location of barycentre of targets, and iii) estimation of the targets ground velocity and direction by image-to-ground transformation.

This preliminary study has shown that the theoretical limit of the estimated velocity accuracy given by the spatial resolution and MS-PAN time lag constraints can be reached using sub-pixel image-to-image warping and accurate target barycentre detection procedures. In particular, with the tested Quickbird sensor, we have estimated a limit of 10Km/h as minimal velocity of targets having the size of a car.

Further efforts will concentrate on testing the different limits associated to different VHR sensors available, and on the development of an automatic procedure able to select the potential moving targets without the manual intervention used in this methodology. The fully automatic target-selection

procedure will be required in case of application of this methodology for detection of unexpected (out of road) or low-velocity moving target in the entire satellite scene, that are difficult to identify by visual inspection.

ACKNOWLEDGEMENTS

The presented work has been supported by the institutional research activity of Action ISFEREA (Information Support for Effective and Rapid External Action) of the European Commission, Joint Research Centre, Institute for Protection of the Citizen (IPSC). The research concept and his implementation have been made possible because of the effective sharing of ideas and tools inside the GMOSS (Global Monitoring for Security and Stability) Network of Excellence.

REFERENCES

- Baltsavias E., Pateraki M., Zhang L., 2001. Radiometric and geometric evaluation of Ikonos geo images and their use for 3D building modeling, Joint ISPRS Workshop "High Resolution Mapping from Space 2001", Hannover, Germany, 19-21 September.
- Dial G., 2003. Measuring Boat Speed, Space Imaging technical paper.
- Dial G., Grodecki J., 2002. Block adjustment with rational polynomial camera models. Proceeding of ASCM-ASPRS Annual Conventions, Washington D.C., April 2002, pp.
- Jacobsen K., 2005. High Resolution Satellite Imaging Systems – Overview, ISPRS Hannover Workshop 2005, High-Resolution Earth Imaging for Geospatial Information, May 17-20, 2005, Hannover, Germany.
- Meyer F., Hinz S., Laika A., Wehling D., and Bamler R., 2006. Performance analysis of the TerraSAR-X Traffic monitoring concept, ISPRS Journal of Photogrammetry and Remote Sensing, 61 (3-4), pp. 225-242
- Munno C.J., Turk H., Wayman J.L., Libert J.M., Tsao T.J., 1993. Automatic video image moving target detection for wide area surveillance, Security Technology Proceedings, Institute of Electrical and Electronics Engineers 1993 International Carnahan Conference on 13-15 Oct. 1993 pp. 47 – 57.
- Paar G., Pözlleitner W., 1992. Robust disparity estimation in terrain modelling for spacecraft navigation. Proc. 11th ICPR, International Association for Pattern Recognition, 1992, pp.
- Reinartz P., Lachaise M., Schmeer E., Krauss T. and Runge H., 2006, Traffic monitoring with serial images from airborne cameras, ISPRS Journal of Photogrammetry and Remote Sensing, 61 (3-4), pp. 149-158.
- Soille P., 2003. Morphological Image Analysis, Springer.
- Suchandt S., Einedera M., Breita H., Runge H., 2006. Analysis of ground moving objects using SRTM/X-SAR data, ISPRS Journal of Photogrammetry and Remote Sensing, 61 (3-4), pp. 209-224.
- Toth C.K. and Grejner-Brzezinska D., 2006, Extracting dynamic spatial data from airborne imaging sensors to support traffic flow estimation, ISPRS Journal of Photogrammetry and Remote Sensing, 61 (3-4), pp. 137-148.

ROAD JUNCTION EXTRACTION FROM HIGH RESOLUTION AERIAL IMAGES

M. Ravanbakhsh, C. Heipke, K. Pakzad

Institute für Photogrammetrie und GeoInformation, Leibniz Universität Hannover
Nienburger Str. 1, D-30167 Hannover, Germany
[ravanbakhsh, heipke, pakzad]@ipi.uni-hannover.de

KEY WORDS: Road junction, Ziplock snake, Gradient Vector Flow (GVF), Geospatial database, High resolution images

ABSTRACT:

Road junctions are important components of a road network. However, they are usually not explicitly modelled in existing road extraction approaches. In this research, we model road junctions in detail as area objects and propose a methodology for their automatic extraction through the use of an existing geospatial database. Prior knowledge derived from the geospatial database is used to facilitate the extraction. We propose a new approach called GVF Ziplock snake that integrates the GVF (Gradient Vector Flow) field as an external force field into a Ziplock snake in order to delineate the junction border. Road extraction results provide fixed boundary conditions for the proposed snake. The approach was tested using Digital Mapping Camera (DMC) images of 0.1 m ground resolution taken from suburban and rural areas. Extraction results are represented in order to illustrate different steps of the method and to prove its feasibility.

1. INTRODUCTION

The need for accurate geospatial databases and their automatic updating is increasing rapidly. Geospatial databases contain various man-made objects among which roads are of special importance as they are used in a variety of applications such as car navigation, transport and fire services. As their extraction from images is costly and time-consuming, automation is seen as a promising solution to this dilemma. However, data acquisition is difficult to automate. The problem for automatic data extraction lies mostly in the complex content of aerial images. To ease the automation of an image interpretation task, prior information can be used (Gerke, 2006), (Boichis et al., 2000), (Boichis et al., 1998), (De Gunst, 1996). This often includes data from an external geospatial database. Road junctions are important components of a road network and if modeled accurately can improve the quality of road network extraction (Boichis et al., 1998). However, there are only few approaches which are dedicated to this task. Junctions are mainly extracted in the context of automatic road extraction. Most of the existing approaches initially concentrate on road extraction to create the road network. Subsequently the extraction of road junctions is realized by perceptual grouping of road hypotheses. In such approaches, junctions are regarded as a point object (Zhang, 2003), (Wiedemann, 2002), (Hinz, 1999). In contrast, in (Gautama et al., 2004) and (Mayer et al., 1998) junctions are treated as planar objects. In (Gautama, 2004) a differential ridge detector in combination with a region growing operator is used to detect junctions and in (Mayer et al., 1998) a snake model is used to delineate junctions. There are also some methods which exclusively deal with junctions. (Barsi et al., 2002) present a road junction operator developed for high-resolution black-and-white images. The operator uses a feed- forward neural network applied to a running window to decide whether it contains a road junction. The drawback of the system is the high level of false alarms. (Wiedemann, 2002) uses a method to improve the quality of the junction extraction which resulted from the approach presented in (Hinz et al., 1999). In (Wiedemann, 2002), several internal and external

evaluation measures are used for the combination of connected roads. The solution with the largest evaluation score is regarded as the solution for the junction.

In (Boichis et al., 2000) and (Boichis et al., 1998) a knowledge based system for the extraction of road junctions is presented. Junctions are classified into four classes and each class is modeled separately. The information derived from an external database is used to generate hypotheses for the junction shape and for main and secondary roads. A drawback of this approach is that image information is not exploited for the formation of main and secondary roads. In (Teoh and Sowmya, 2000) junctions are recognized by using several rules and a supervised learning method. They are classified into several types and each type is treated individually. Rules address several attributes of the junctions. The range of values for the attributes comes from a large set of data. Since the result of the approach is not available in publications, its performance cannot be assessed.

Road junctions in road network extraction systems have mainly been modeled as point objects at which three or more road segments meet (Zhang, 2003), (Wiedemann, 2002), (Hinz, 1999). The junction position in such systems is computed by simple extension of neighboring road segments. This kind of modeling does not always reflect the required degree of detail. In Fig.1, vector data is superimposed on a sample image to describe the problem. In the given image resolution, the junction centre covers an area, so it should be considered as an area object. Thus, a detailed modeling of junctions is needed for data acquisition purposes in large scales. In this paper, junctions are modeled in detail as area objects. The approach uses an existing geospatial database and leads to the extraction of refined road junction data. In section 2, a short overview of the proposed strategy is given. In section 3 junctions are classified and modeled. Various steps of the proposed strategy are described in section 4. Some results of the road junction extraction are shown and discussed in section 5. In section 6, an outlook for future research is given.



Figure 1. Superimposition of vector data on a high resolution aerial image

2. OVERVIEW OF THE APPROACH

We used Digital Mapping Camera (DMC) images with a ground resolution of 0.1 m. The images were taken over rural and suburban areas. Besides the image, the German ATKIS (Amtliches Topographisch-Kartographisches Information System) with a content approximately equivalent to that of 1:25000 scale topographic maps is entered into the system as part of the input data. In ATKIS, the planimetric accuracy for a road object is defined as ± 3 m.

Three main parts of our approach are introduced (Fig. 2):

- Pre-analysis of geospatial database
- Extraction of road arms
- Road junction reconstruction

In the first part, the geospatial database is analyzed resulting in a rough idea of the junction appearance in the image. In the second part, road arms are extracted within a limited area around the initial junction position with a size of 70×70 m² (700×700 pixels). A road arm is the longest straight road segment connected to the junction center. In the third part, the road junction is reconstructed using a snake-based approach that integrates the Gradient Vector Flow (GVF) (Xu and Prince, 1997) as an external force field into a Ziplock snake

(Neuenschwander et al., 1997) in order to delineate the junction border.

3. MODEL

3.1 Road junction classification

For traffic safety, different kinds of junctions have been designed. Different types of junctions have different properties and construction specifications. In our work, junctions have been classified with an emphasis on the most common features among them. We classified junctions into four main classes: simple, complex, roundabout and motorway. Simple junctions contain three or more road arms without islands in the center. In contrast, complex junctions do contain islands in the center. A road model for simple junctions complies with the classic road model in which roads are defined as quadrilateral objects with parallel edges and a constant width. In three other classes, roads do not necessarily follow this classic model because there might be, for instance, roads with a changing width. The main difference between motorway junctions and others is that crossing roads are not at the same height. In other words, roundabouts, simple and complex junctions are defined in two dimensional space but motorway junctions are defined in three dimensional space. In complex junctions, connected roads enclose an area in which islands are located despite the roundabout whose roads intersect its large circular island. In this paper, we focus on simple junctions only.

3.2 Road junction model

The conceptual model of simple road junctions is represented and described in Fig. 3. According to the model, a junction area is composed of two parts: the junction itself and the road arms. The junction, where road arms are connected, is composed of the junction border at its central area. A road arm is defined in terms of geometry and radiometry as following:

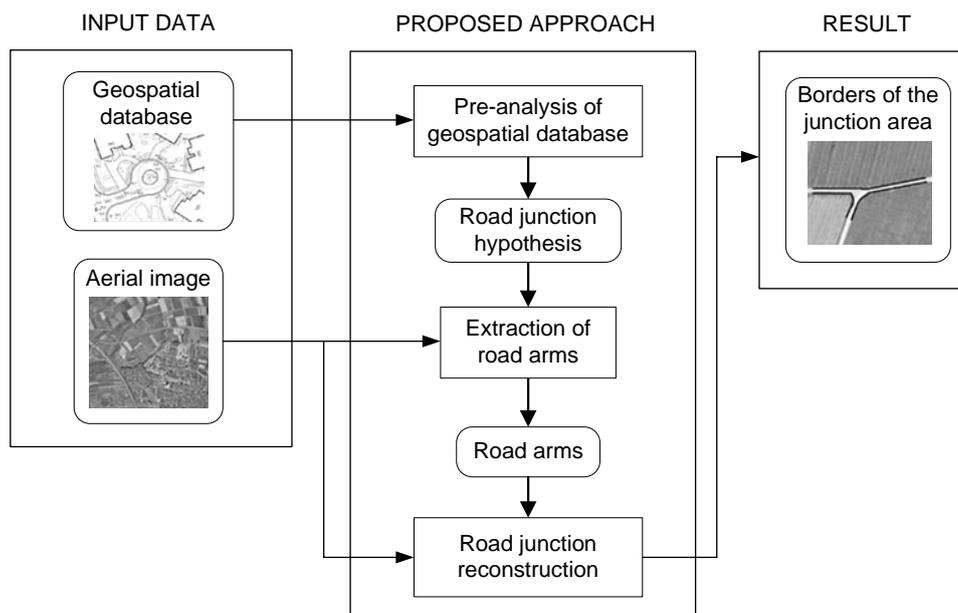


Figure 2. Proposed system organisation

- Geometry: A road arm is a rectilinear object which is represented as a ribbon with a constant width and two parallel road edges.
- Radiometry: A road arm is considered to be a homogeneous region with high contrast to its surroundings. The absolute brightness depends on the surface material.

Disturbances such as occlusions and shadows are not explicitly included in the model at this stage.

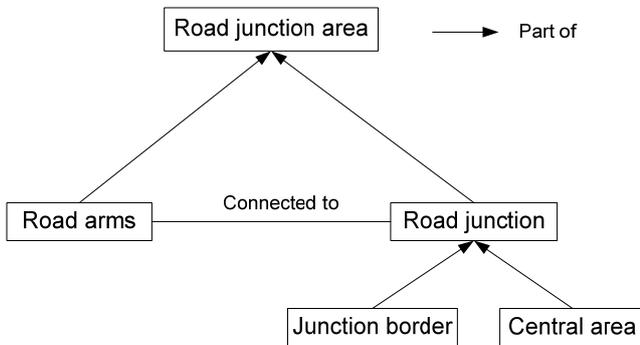


Figure 3. Road junction model

4. EXTRACTION STRATEGY

The strategy guides the system where and how to extract objects in image space and how to interpret extracted elements (De Gunst, 1996) (Fig. 2). In this work, we make use of prior knowledge derived from a geospatial database to extract only those parts of the object which are consistent with the corresponding content of the geospatial database. As mentioned above our strategy consists of three steps.

4.1 Pre-analysis of geospatial data base

The geospatial database we used contains explicit geometric and implicit topologic information about road junctions. Topologic information determines the number of roads connected to the junction centre and geometric information provides us with the approximate location of the junction and the width of the connected roads. A road junction in the geospatial database is composed of a centre point at which a

few lines or polylines meet. Using vector coordinates of lines, we compute road directions. In the road extraction step, geometric and topologic information are used to construct road segments.

4.2 Extraction of road arms

Roads can be bent in different ways, for instance, in a simple curved form, serpentine curve or in a state with changing width. However, in the area close to the junction centre they are mainly straight because of traffic safety regulations. This fact leads us to extract long and straight road segments near the junction centre. They are called road arms (Fig. 4). To illustrate various steps and show the result of each one, we used some image samples. In order to apply the geometric part of the road model, edges are extracted from the image using the Deriche edge detector. Subsequently, a thinning operation is applied, yielding one pixel wide edges. The edges are approximated by polygons to facilitate further processing. We call the result of this step edge segments. Edge segment parameters like image coordinates of endpoints, length and direction are computed. These parameters are used later for the road segment construction. We then group edge segments based on the direction of connected roads from the geospatial database. The number of groups corresponds to the number of connected roads (Fig. 5-a). Since road directions derived from the geospatial database are regarded as reference directions, each group should contain parallel edge segments having a direction similar to their reference direction. It is noted that the direction difference between edge segments in each group must be below a predefined threshold, 15 degrees. Next, as a part of the geometric road model, width and the width constancy between two edge segment candidates are checked. A prerequisite for this step is that two candidates must overlap and must have opposite directions (anti-parallelism condition). Next, radiometric properties of the resulting road segments are investigated. According to the radiometric road model, the area between two road edges should be bright and homogeneous. This means the mean gray value within each area must fall into a predefined range and its variance must be smaller than a predefined threshold. Resulting road segments are verified by extracting road centerlines from an image of reduced resolution (Heipke et al., 1995) (Fig. 5-b).

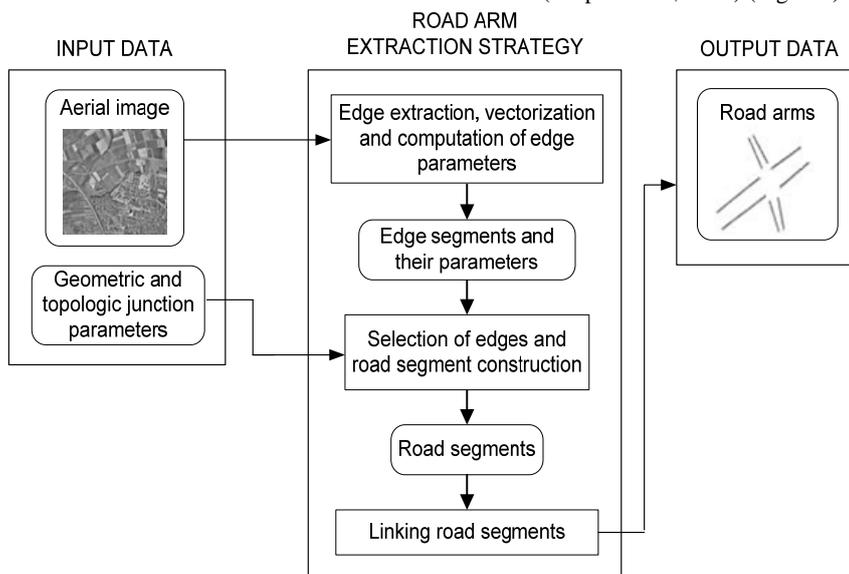


Figure 4. Road arm extraction

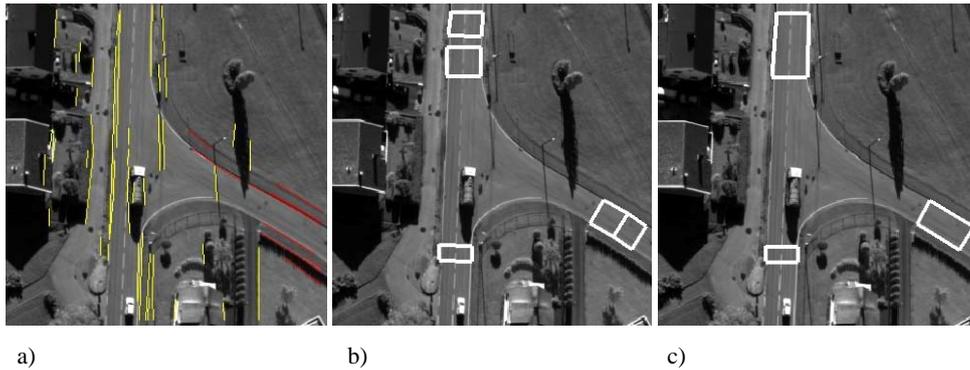


Figure 5. (a) illustrates two groups of edge segments in red and yellow. Yellow edge segments contain two groups since related roads are collinear. (b) shows the resulting road segments. (c) shows the extracted road arms.

Although we have so far extracted groups of road segments, what is needed is one road arm per group. To achieve this purpose, road segments within each group are linked (Fig. 5-c). Next, the orientation of each resulting road arm is investigated with the aim to decide which end point faces the junction. This information could be taken from the geospatial data base, but in order to be more independent of the geometric accuracy of the database, we compute the distance between the endpoints of pairs of road arms, and subsequently select the endpoints sharing the shortest distance.

4.3 Road junction reconstruction

4.3.1 Ziplock snake

Traditional snakes (Kass et al., 1987), or parametric active contours are polygonal curves with which is associated an objective function that combines an image term, external energy, measuring the edge strength and a regularization term, internal energy, minimizing the tension and curvature. The curve is deformed so as to optimize the objective function and, as a result, to extract the image edges. Traditional snakes are sensitive to noise and need a close initialization. Since the junction border is composed of open curves with various degrees of curvature, close initialization often cannot be provided. As a result, traditional snakes get stuck in an undesirable local minimum. To overcome these limitations, the Ziplock snake model was proposed (Neuenschwander et al., 1997). Ziplock snakes are now briefly described here to provide a basis to introduce our approach. A Ziplock snake consists of two parts: an active part and a passive part (Fig. 6). The two parts are separated by moving force boundaries, and the active part is further divided into two segments, indicated as head and tail respectively. The initial positions of the head and tail segments are specified by an operator or a preprocessing module. Unlike the procedure for a traditional snake, the external force derived from the image is turned on only for the active parts. Thus the movement of passive vertices is not affected by any image forces. Starting from two short pieces, the active part is iteratively optimized to image features, and the force boundaries are progressively moved toward the center of the snake. Each time that the force boundaries are moved, the passive part is re-interpolated using the position and direction of the end vertices of the two active segments. Optimization is stopped when force boundaries meet each other. We call the external force used in the Ziplock snake traditional force field.

Ziplock snakes need far less initialization effort and are less affected by the shrinking effect from the internal energy term. Furthermore, the computation process is more robust because

the active part whose energy is minimized is always quite close to the contour being extracted. A Ziplock snake is an appropriate method to delineate the junction border if accurate initial points could be provided. We assume that these accurate fixed points can be obtained from the road extraction result.

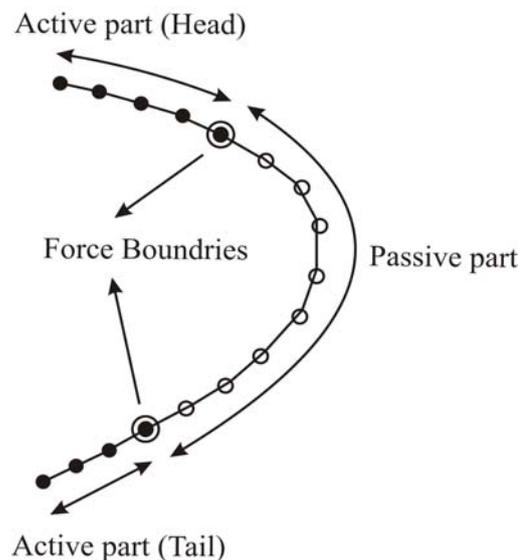


Figure 6. Illustration of a ziplock snake during optimization. A Ziplock snake, fixed at head and tail, consists of two parts, the active and the passive vertices. These areas are separated by moving force boundaries. The active parts of the snake consist of head and tail segments.

Nevertheless, Ziplock snakes are easily confused by disturbances caused by various features such as trees and their shadows, different kinds of road markings and the shadows from buildings, traffic lights, power lines and traffic signs. A strong internal energy can considerably decrease the effect of disturbing features. Furthermore, a global initialization is introduced to assure that snake vertices are distributed across the entire object boundary. The global initialization is provided by pairs of lines, each of which is defined by close endpoints and the intersection point of their related road sides (Fig. 7). These lines are divided up into equidistance vertices surrounding curve parts of the road junction.



Figure 7. Illustration of the global initialization. Black lines are defined by two close endpoints and the intersection point of their respective road sides.

We tested the Ziplock snake method on several samples (Fig. 8-a, b). As opposed to low-resolution images where the road surface is quite homogeneous and road markings have less disturbing effects, in high-resolution images there are various disturbing features destabilizing Ziplock's active parts and subsequently hinder their motion. Furthermore, disturbing features, such as road markings as are shown in Fig. 8-a, besides destabilizing the active parts, i.e. no convergence occurs, pull the contour away from the desired boundaries. However, if the external force has a large capture range, it draws the contour back toward the object boundaries even from far distances. Since a traditional force field has a small capture range, Gradient Vector Flow (GVF) as an alternative external field was used.



Figure 8: Ziplock snake optimization using tradition force field as an external force. Black curves are the result of snake optimization. White lines are the extracted road arms. Since the capture range of the external force field is small, the iterations are stopped near the initial positions.

4.3.2 GVF Ziplock snake

The GVF field (Xu and Prince, 1997) was proposed to address two issues: a poor convergence to concave regions and problems associated with the initialisation. It is computed as a spatial diffusion of the gradient of an edge map derived from the image. This computation causes diffuse forces to exist far from the object, and crisp force vectors near the edges. The GVF field points toward the object boundary when very near to the boundary, but varies smoothly over homogeneous image regions, extending to the image border (Fig. 9). The main advantage of the GVF field is that it can capture a snake from a long range. Thus, the problem of far initialization can be alleviated.

The GVF is defined to be the vector field $G(x, y) = (u(x, y), v(x, y))$ that minimizes the energy functional:

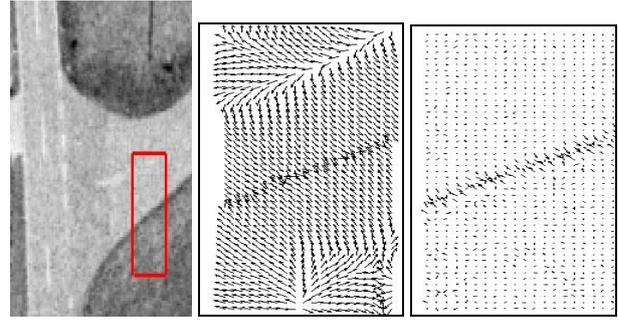


Figure 9: a) a small part of a test image b) GVF field c) Tradition force field

$$E = \iint \mu (u_x^2 + u_y^2 + v_x^2 + v_y^2) + |\nabla f|^2 |G - \nabla f|^2 dx dy$$

Where $f(x, y)$ is derived from the image having the property that it is larger near the image edges. μ is a regularization parameter which should be set according to the amount of noise present in the image. Using the *calculus of variations*, it can be shown that the GVF can be found by solving the following Euler equations:

$$\mu \nabla^2 u - (u - f_x)(f_x^2 + f_y^2) = 0$$

∇^2 is the Laplacian operator.

We call the Ziplock snake that uses the GVF field as its external force a GVF Ziplock snake.

Let $V(s) = (x(s), y(s))$ be a parametric active contour in which S is the curve length and x and y are the image coordinates of the 2D-curve. The internal snake energy is then defined as

$$E_{int}(V(s)) = \frac{1}{2} (\alpha(s) |V_s(s)|^2 + \beta(s) |V_{ss}(s)|^2)$$

Where V_s and V_{ss} are the first and second derivative of V in respect to S . The functions $\alpha(s)$ and $\beta(s)$ control the elasticity (the first term) and the rigidity (the second term) of the contour respectively.

The formulation for GVF Ziplock snake's motion can be written in the form (Kass et al., 1987):

$$V [t] = (K + \gamma I)^{-1} * (\gamma V [t-1] - \kappa P_v |_{V[t-1]})$$

Where γ stands for the viscosity factor (step size) determining the rate of converges and t is the iteration index. κ alters the strength of the external force. The pentadiagonal matrix K contains the internal energy functions (α, β) and P_v is the GVF external force field.

4.3.3 Implemetation issues

In our implementation, we chose $\alpha = 0.001$ because the larger value forces the snake to become and then stay straight. Based on many tests on junctions of different shapes $\beta = 5$ was chosen to let the contour become smooth. To reduce the effect of disturbances, we set $\kappa = 0.1$, i.e. giving more weight to the internal force than to the external. Snake spacing refers to the distance between the sampled snake vertices. By experiment, denser snake vertices are more likely to be trapped and

destabilized by small disturbances (Fig. 10-a). Thus, convergence, assuming a small threshold to be achieved for the displacement from one iteration to the next, cannot be reached. Conversely, when snake vertices are too sparse, object boundaries with high change of curvature, as it is the case in road junctions, cannot be accurately delineated (Fig. 10-b). The snake spacing was confined within the range 7-10 pixels depending on how far two adjacent road arms are, so that their movement is less likely to be blocked by trees, single cars and road markings such as stop lines and cross walks.



a)

b)

Figure 10. In (a), snake spacing is 2 pixels, which signifies dense vertices, and the accuracy of convergence is 1/10 pixel. Active contours are caught in the shadow cast by the truck. In (b), snake spacing is 12 pixels and the accuracy of convergence is 1/200 pixel. The optimized active contour (black line) cannot delineate the junction border.

Likewise, the disturbing effect of the shadow cast by traffic lights and power lines are easily resolved. However, rows of cars standing behind the traffic light, and every large body of shadow from a row of buildings or trees might mislead the snake to converge to wrong boundaries, because in such cases, a large number of vertices are trapped in the disturbing object. This situation affects the motion of neighbouring vertices and eventually results in the delineation of the disturbing object boundaries or the leakage into either the road junction area or the surroundings. In our proposed approach, the force boundary is advanced one vertex per iteration when we can verify that the motion of the corresponding active part has stabilized. We evaluate each active part individually by testing if the displacement in x and y direction are less than a predefined threshold (1/200 pixel). Once two force boundaries collide, the viscosity factor is increased and the optimization is repeated simultaneously on all vertices since it improves the quality of the final result at places where there is a small deviation from the object boundaries. Intuitively, a little value of γ is chosen when the process starts, (because the curve is far from the solution) and a larger value when the snake is close to the contour (Berger and Mohr, 1990).

5. RESULTS

The contour-following property of Ziplock snakes, when combined with the large capture-range capability of the GVF force field, enables delineation of complex man-made objects in high-resolution aerial imagery. To demonstrate this potential, we tested the proposed approach on many road junction samples with the same set of parameters. Some of them are shown (Fig. 11). The problem illustrated in Fig. 8 is fixed. Furthermore, the approach can deal with disturbances caused by road markings and the shadow (bottom row image samples).



Figure 11. Top row image samples show rural areas and those in the bottom row are from suburb areas. In both groups, disturbances caused by road markings are resolved. Furthermore, the results in suburb areas are accurate even in the presence of shadows.

6. OUTLOOK

In this paper, we presented a new snake-based approach called GVF Ziplock snake to delineate the junction border resulting in the extraction of simple road junctions in the context of rural and suburb areas through the use of an existing geospatial database. We found that prior knowledge from an existing geospatial database can considerably facilitate the extraction. Furthermore, we demonstrated that our approach can overcome various kinds of disturbances. Further investigation into the possibility of using Balloon snakes in junctions whose central area lacks sufficient contrast with the surroundings causing the snake to be drawn outward of the junction area and to become straight rather than curvy, are warranted. Investigations into modelling islands' properties and extracting complex junctions and roundabouts are our next goals. Furthermore, optimal selection of the snake spacing constraint and the convergence threshold in images with different ground resolutions is desirable.

REFERENCES

- Baltsavias, E.P., 2004. Object extraction and revision by image analysis using existing geodata and knowledge: current status and steps towards operational systems. In: *IntArchPhRS.*, Vol. 58(3-4), pp. 129-151.
- Barsi, A., Heipke, C., Willrich, F., 2002. Junction Extraction by Artificial Neural Network System – JEANS. In: *IntArchPhRS.*, Vol. 34, Part 3B, pp. 18-21.
- Berger, M.O., Mohr, R., 1990. Towards autonomy in active contour models. In: *The tenth International Conference on Pattern Recognition*, Atlantic City, NJ, pp. 847-851.
- Boichis, N., Cocquerez, J.-P., Airault, S., 1998. A top down strategy for simple crossroads extraction. In: *IntArchPhRS.*, Vol. XXXII, part 2/1, pp. 19-26.
- Boichis, N., Viglino, J.-M., Cocquerez, J.-P., 2000. Knowledge based system for the automatic extraction of road intersections from aerial images. In: *IntArchPhRS.*, Vol. XXXIII, Supplement B3, pp. 27-34.
- Gautama, S., Goeman, W., D'Haeyer, J., 2004. Robust detection of road junctions in VHR images using an improved ridge detector. In: *IntArchPhRS.*, Vol. XXXV, Part B3, Istanbul, pp. 815-819.
- Gerke, M., 2006. Automatic Quality Assessment of Road Databases Using Remotely Sensed Imagery. *Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Universität Hannover*, No. 261; also in: *Deutsche Geodätische Kommission, Reihe C*, No. 599, 105 p.
- De Gunst, M., 1996. Knowledge-based interpretation of aerial images for updating of road maps. Ph.D. thesis, Delft University of Technology, the Netherlands.
- Heipke, C., Steger, C., Multhammer, R., 1995. A hierarchical approach to automatic road extraction from aerial imagery. In: *McKeown D.M., Dowman I., (Eds.), Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision II*, SPIE Proceedings (2486), 222-231.
- Hinz, S., Baumgartner, A., Steger, C., Mayer, H., Eckstein, W., Ebner, H., Radig, B., 1999. Road extraction in rural and urban areas. In: *Semantic Modeling for the Acquisition of Topographic Information from Images and Maps*. In: *SM-ATI'99*. München, pp. 7-27.
- Kass, M., Witkin, A., Terzopoulos, D., 1987. Snakes: Active contour models. In: *Int. J. Computer Vision*, 1(4):321-331.
- Mayer, H., Laptev, I., Baumgartner, A., 1998. Multi-Scale and Snakes for Automatic Road Extraction. In: *Proc. Fifth European Conference on Computer Vision*, Freiburg, Germany, Vol. 1406 of Springer Verlag Lecture Notes in Computer Science, pp.720-733.
- Neuenschwander, W. M., Fun, P., Iverson, L., Szekely, G., Kubler, O., 1997. Ziplock snakes. In: *Int. J. Comput. Vis.*, vol. 25, no. 3, pp. 191-201.
- Teoh, C., Sowmya, A., 2000. Junction extraction from high resolution images by composite learning. In: *IntArchPhRS.*, Vol. XXXIII, part B3, pp. 882-888.
- Wiedemann, C., 1999. Completion of automatically extracted road networks based on the function of roads. In: *IntArchPhRS.* Vol. 32(3-2W5), pp. 209-215.
- Wiedemann C., 2002. Improvement of Road Crossing Extraction and External Evaluation of the Extraction Results. In: *IntArchPhRS.*, Vol. 34, Part 3B, pp. 297-300.
- Xu, C., Prince, J., 1998. Snakes, shapes, and gradient vector flow. In: *IEEE Trans. Imag. Proc.*, vol. 7, pp. 359-369.
- Zhang, C., 2003. Updating of Cartographic Road Databases by Images Analysis. Ph.D. thesis, Institute of Geodesy and Photogrammetry, ETH Zurich, Switzerland, IGP Mitteilungen Nr. 79.

SINGLE TREE DETECTION IN FOREST AREAS WITH HIGH-DENSITY LIDAR DATA

J. Reitberger^{a,*}, M. Heurich^b, P. Krzystek^a, U. Stilla^c

^a Dept. of Geoinformatics, Munich University of Applied Sciences, 80333 Munich, Germany
(reitberger, krzystek)@fhm.edu

^b Dept. of Research, Bavarian Forest National Park, 94481 Grafenau, Germany
Marco.Heurich@npv-bw.bayern.de

^c Photogrammetry and Remote Sensing, Technische Universitaet Muenchen, 80290 Munich, Germany
stilla@tum.de

Commission III, WG III /3

KEY WORDS: LIDAR, analysis, segmentation, forestry, vegetation

ABSTRACT:

The study presents a novel method for delineation of tree crowns and detection of stem positions of single trees from dense airborne LIDAR data. The core module of the method is a surface reconstruction that robustly interpolates the canopy height model (CHM) from the LIDAR data. Tree segments are found by applying the watershed algorithm to the CHM. Possible stem positions of the tallest trees in the segments are subsequently derived from the local maxima of the CHM. Additional stem positions in the segments are found in a 3-step algorithm. First, all the points between the ground and the crown base height are separated. Second, possible stem points are found by hierarchical clustering these points using their horizontal distances. Third, the stem position is estimated with a robust RANSAC-based adjustment of the stem points. We applied the method to small-footprint full waveform data that have been acquired in the Bavarian Forest National Park with a point density of approximately 25 points per m². The results indicate that the detection rate for coniferous trees is 61 % and for deciduous trees 44 %, respectively. 7 % of the detected trees are false positives. The mean positioning error is 0.92 cm, whereas the additional stem detection improves the position on average by 22 cm.

1. INTRODUCTION

Laser scanning has become the leading edge technology for the acquisition of topographic data and the mapping of the Earth's surface and offers several advantages for forest applications. The laser beam may penetrate the forest structure, and the technique provides 3D information at a high point density and intensity values at a specific wavelength. Since over a decade, conventional LIDAR - recording the first and last pulse - has been widely used to successfully retrieve forest parameters like tree height, crown diameter, number of stems, stem diameter and basal area (Hyyppä et al., 2004). Recently, tree species classification could be tackled with first/last-pulse scanning systems providing high point density (Holmgren et al., 2004; Heurich, 2006; Brandtberg, 2007). Moreover, the novel airborne full waveform technology promising new possibilities has been lately applied to tree species classification (Reitberger et al., 2006).

Approaches to tree species classification are usually based on a single tree segmentation that delineates the tree crown from the outer geometry of the forest surface. The methods have in common to reconstruct – at least locally – the canopy height model (CHM) to find the local maximum in the CHM as the best guess for the stem position and to delineate a segment polygon as the tree crown. For instance, Hyyppä et al. (2001) interpolate a local CHM from the highest laser reflections, Persson et al. (2002) apply the active contour algorithm, and Solberg et al. (2006) interpolate the CHM with a special gridding method and subsequently smooth the CHM with an appropriate Gaussian filter. Stem positions are derived from the interpolated CHM at the highest positions (Solberg et al., 2006) or from a special local tree shape reconstruction (Brandtberg, 2007). Tree crowns are typically derived with the watershed

algorithm (Pyysalo et al., 2002) or by a slope-based segmentation (Persson et al., 2002; Hyyppä et al., 2001). Recently, Solberg et al. (2006) proposed a region growing method that starts from local surface maximums and finds crown polygons optimised in shape.

The drawback of these methods is that they are solely oriented on the CHM. The CHM is reconstructed from the raw data in an interpolation process that smoothes the data to some extent. The degree of smoothing affects directly the success rate in terms of false positives and negatives. Moreover, in some cases neighbouring trees do not appear as two clear local maximums. Thus, approaches that solely use the CHM will be restricted in the success rate anyway, especially in heterogeneous forest types where group of trees grow close together. So far, little attention has been paid to reconstruct trees using information like laser hits on the stem, mainly because of the low spatial point density. Detected tree stems could be used to improve the CHM-based segmentation of the single trees in terms of the detection rate of trees and the position of the trees. Moreover, new full waveform systems have the potential to detect significantly more reflections in the tree crown and hence highly resolve the internal tree structure.

The objective of this paper is (i) to present a novel method that segments single trees with a robust surface reconstruction method in combination with the watershed algorithm using full waveform LIDAR data, (ii) to introduce a novel approach to stem detection that clusters hierarchically potential stem reflections and reconstructs the stem with a RANSAC-based adjustment, and (iii) to show how the detection rate and position of single trees is improved.

The paper is divided into five sections. Section 2 focuses on the segmentation of the single trees and the reconstruction of the

* Corresponding author.

tree stems. Section 3 shows the results which have been obtained from full waveform data acquired in May 2006 by the Riegl LMS-Q560 system in the Bavarian Forest National Park. Finally, the results are discussed with conclusions in section 4 and 5.

2. METHODOLOGY

2.1 Decomposition of full waveform data

Let us assume that full waveform LIDAR data have been captured in a region of interest (ROI). The waveforms are decomposed by fitting a series of Gaussian pulses to the waveforms (Figure 1). Thus, for each reflection i the vector $\mathbf{X}_i^T = (x_i, y_i, z_i, W_i, I_i)$ ($i=1, \dots, N_{ROI}$) is provided with (x_i, y_i, z_i) as the coordinates, W_i as the pulse width and I_i as the intensity of the reflection (Reitberger et al., 2006; Jutzi and Stilla, 2005). Note that basically each reflection can be detected by the waveform decomposition. This is remarkable since conventional LIDAR systems – recording at most five reflections – have a dead zone of about 3 m which make these systems effectively blind after a single reflection.

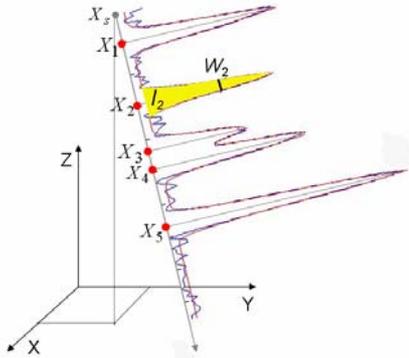


Figure 1. 3D points and attributes derived from a waveform

2.2 Segmentation

The segmentation of the tree crowns is achieved by deriving the CHM from 3D points which are best representing the outer tree crown geometry. The ROI is subdivided into a grid having a cell spacing of cp and N_R cells (Figure 2). Within each cell of size cp^2 , the highest 3D point is extracted and corrected with respect to the ground level z_j^{ground} , i.e.

$z_j^{CHM} = z_j - z_j^{ground}$ ($j=1, \dots, N_R$). The ground level z_j^{ground} is estimated from a given digital terrain model (DTM) by bilinear interpolation. In the next step, all the highest 3D points $\mathbf{X}_j^T = (x_j, y_j, z_j^{CHM})$ ($j=1, \dots, N_R$) of all N_R cells are robustly interpolated in a grid that has N_X and N_Y grid lines and a grid width gw . The special adjustment approach (Krzystek et al., 1992) interpolates the $N_{CHM} = N_X * N_Y$ grid points

$\mathbf{X}_{int_j}^{CHM T} = (x_{int_j}^{CHM}, y_{int_j}^{CHM}, z_{int_j}^{CHM})$ ($j=1, \dots, N_{CHM}$) and filters the 3D points \mathbf{X}_j in a 2-phase iterative Gauß-Markoff process. Thanks to constraints on the curvature and torsion of the surface, the interpolation smoothes and regularises the surface in case of an ill-posed local situation. Moreover, single outlying 3D points, which do not represent the CHM surface, are down-weighted by weighting functions in dependence on the distance of the 3D points to the surface. This iterative adjustment scheme works like an edge preserving filter that discards outliers, closes gaps in the surface if no 3D points could be

found in the cells, and preserves surface discontinuities. The result is a smoothed CHM having N_{CHM} equally spaced posts. Finally, the tree segments are found by applying the watershed algorithm (Vincent and Soille, 1991) to the CHM. The local maximums of the segments define the N_{seg} tree positions $(X_{stem_i}^{CHM}, Y_{stem_i}^{CHM})$ ($i=1, \dots, N_{seg}$).

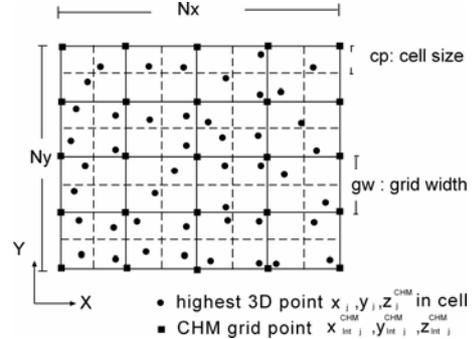


Figure 2. Finite element interpolation for a patch of 16 grid meshes

2.3 Stem detection

Tree stems in the individual tree segments are detected in a 3-step procedure.

Step 1: The N_S points X_j^{Seg} ($j=1, \dots, N_S$) within a tree segment are cleared from ground points by discarding all points within a given height bound $Z_{threshold} = 1$ m to the DTM.

Step 2: The goal of the second step is to derive the crown base height h_{base} of the tree in order to subdivide the tree into the stem area and the remaining crown area. This coarse tree subdivision is achieved by (i) splitting the tree into l layers with height of 0.5 m (Figure 3a), (ii) calculating the number of points n_i per layer, (iii) forming the vector $N_p = \{n_i / N_S\}$ ($i=1, \dots, l$), (iv) smoothing N_p with a 3x1 Gaussian filter and, finally, (v) defining h_{base} as the height that corresponds to 0.15 % of the total number of points per segment (Figure 3b). All the N_{stem} points below h_{base} are potential stem points. Note that the remaining points can result from one or even several stems or from the understorey. The following hierarchical clustering scheme is applied to these points after calculating the Euclidian

distance matrix $D_{stem} = \{d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}\}$; $i=1, \dots, N_{stem}; j=1, \dots, N_{stem}; i \neq j$ (Heijden et al., 2004).

1. Assign each point to its own cluster, resulting in N_{stem} clusters.
2. Find the closest pair of clusters and merge them in to one cluster. The number of clusters reduces by one.
3. Compute the distance d between the new clusters and each of the old clusters.
4. Repeat steps 2 and 3 until all items are clustered into a single cluster of size N_{stem} or a predefined number of clusters is achieved.

In this clustering process the distance between two clusters C_i and C_j is defined as the shortest distance from any point in one cluster to any point in the other cluster. The clustering yields a dendrogram which shows at which distance the clusters are grouped together. By defining a minimum distance $d_{min} = 1.2$ m

between the cluster centres the most suitable number of clusters $N_{cluster}$ is selected.

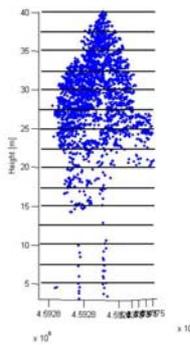


Figure 3a. Tree layers

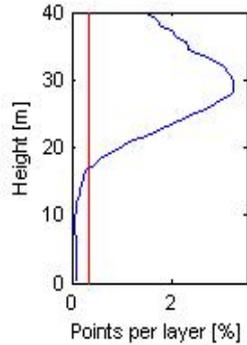


Figure 3b. Base height of a tree

Step 3: The final finding of the stems is achieved by applying a RANSAC-based 3D line adjustment to all the $N_{cluster}$ clusters and labelling all 3D lines with an incident angle smaller than $\alpha = 7^\circ$ and a minimum number of 3 points as stems g_{stem} . This robust procedure eliminates clusters that result from the understorey and do not show a vertical main direction. Also, it cleans the cluster point cloud from erroneous points. The stem positions $(x_{stem}^{det}, y_{stem}^{det})$ are calculated as the intersection of the stem g_{stem} with the DTM $(x_{stem}^{det}, y_{stem}^{det}) = \{DTM \cap g_{stem}\}$. Finally, the height of the tree top h_{tree} is derived from the highest laser point that lies within the cylinder V_{stem} . The cylinder V_{stem} is defined by the 3D line g_{stem} as the centre line of the cylinder and the radius $R = 1$ m. Note that several stems can be found within a tree segment. However, the tree height is dependent on the highest raw data within the stem cylinder V_{stem} .

3. EXPERIMENTS

3.1 Material

Experiments were conducted in the Bavarian Forest National Park which is located in south-eastern Germany along the border to the Czech Republic (49° 3' 19" N, 13° 12' 9" E). There are four major test sites of size between 591 ha and 954 ha containing alpine spruce forests, mixed mountain forests and spruce forests as the three major forest types. 11 sample plots with an area size between 1000 m² and 3600 m² were selected in the mixed mountain forests dominated by Norway spruce (*Picea abies*) and European beech (*Fagus sylvatica*). Some fir trees (*Abies alba*), Sycamore maples (*Acer pseudoplatanus*) and Norway maples (*Acer platanoides*) also occur in the sample plots. The height above sea level varies between 610 m and 770 m. Reference data for all trees with diameter at breast height (DBH) larger than 10 cm have been collected in May 2006 for 438 Norway spruces, 477 European beeches, 74 fir trees, 20 Sycamore maples and three Norway maples. Several tree parameters like the DBH, total tree height, stem position and tree species were measured and determined with the help of GPS, tacheometry and the 'Vertex' III system. Moreover, a DTM with a grid size of 1 m and an absolute accuracy of 25 cm was available for all the test sites. It was generated from LIDAR data which had been acquired in 2003 (Heurich, 2006). For two of the four test sites, which contain the 11 sample plots, full waveform data have been collected by Milan Flug GmbH with the Riegl LMS-Q560 scanner. The Riegl scanner was flown in May 2006 after snowmelt but prior to foliation. The vertical sampling distance was 15 cm because of the scanner's

sampling rate of 1 GHz. The pulse width at half maximum reached 4 ns which is equivalent to 60 cm. The size of the footprint of 20 cm was caused by the beam divergence of 0.5 mrad and the flying altitude of 400 m. Finally, the average point density was 25 points/m². Thorough calibration of the bore sight alignment and strip adjustment was carried out prior to the data evaluation. Controlled tests showed that the strips had been adjusted with a standard deviation of 15 cm in planimetry and 10 cm in height. The procedures for segmentation and subsequent stem detection were applied to all the plots in a batch procedure without any manual interaction.

3.2 Segmentation

Figures 4 and 5 show a typical sample area containing several coniferous trees. The tree tops derived from the local maximums of the CHM correspond in some cases with the reference trees reasonably. However, some tree tops are deviating considerably from the true position. Moreover, some segments contain more than one reference tree although only one tree top has been derived from the CHM. The main reasons are that (i) a group of trees form locally a well-defined maximum and (ii) the surface reconstruction smoothes the surface too much so that neighbouring trees cannot be isolated. In both cases the single trees are not detected and hence the segment represents a group of trees rather than a single tree.

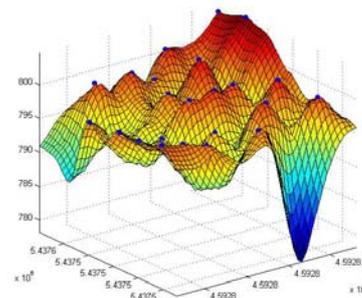


Figure 4. Reconstructed CHM with local maximums as tree tops

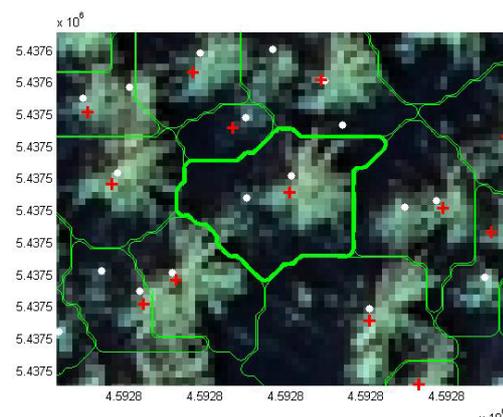


Figure 5. Orthophoto with sample segment containing two reference coniferous trees (=white dots) and one local maximum (=red cross)

3.3 Stem detection

The stem detection takes advantage of additional high-density point information the waveform decomposition provides underneath the CHM. In case that only sparse understorey is below the base height stem points are successfully detected by

the hierarchical clustering and the RANSAC-based stem reconstruction. Figures 6a and 6b show the stem points for the sample segment in figure 5 found by the clustering scheme given in section 2.3. The two stems are clearly isolated by applying the angle constraint of 7^0 to the stems approximated with RANSAC. Moreover, the single stem position derived from the CHM maximum is significantly improved by the new stem position (Figure 7). Also, some neighbouring segments show similar results. For example, in the segment directly above the central segment even 3 stems could be detected whose positions correspond fairly well with the reference trees. Thus, the stem detection provides additional single trees that cannot be found solely using the CHM information and improves the position of trees derived from the CHM maximum.

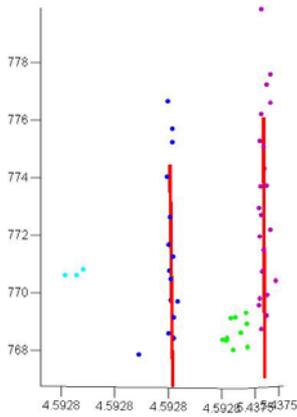


Figure 6a. Stem point clusters and stems reconstructed with RANSAC

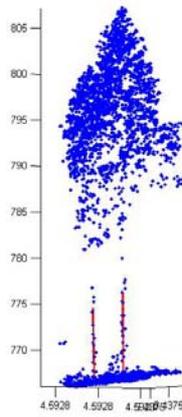


Figure 6b. The neighbouring trees and the reconstructed stems

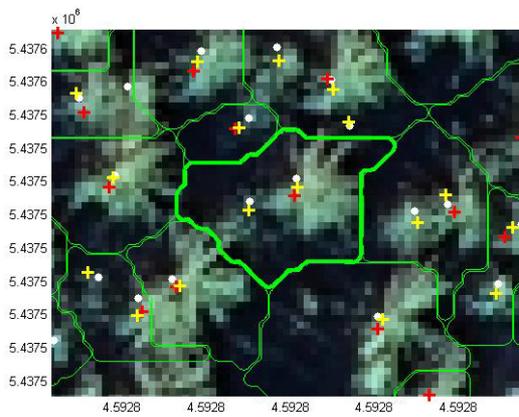


Figure 7. Sample segment with two reference coniferous trees (=white dots), two detected stems (=yellow crosses) and the local CHM maximum (=red cross)

3.4 Evaluation

The tree positions $(x_{stem}^{CHM}, y_{stem}^{CHM})$ from the segmentation and the tree positions $(x_{stem}^{det}, y_{stem}^{det})$ from the stem detection are compared with reference trees if (i) the distances to the reference trees is smaller than 60 % of the mean tree distance of the plot and (ii) the height difference between h_{tree} and the height of the reference tree is smaller than 15 % of the top height h_{top} of the plot, where h_{top} is defined as the average

height of the 100 highest trees per ha (Heurich, 2006). If a reference tree is assigned to more than one tree position, the tree position with the minimum distance to the reference tree is selected. Reference trees that are linked to one tree position are so-called “detected trees” and reference trees without any link to a tree position are treated as “non-detected” trees. Finally, a tree position without a link to a reference tree results as a “false detected” tree.

3.5 Results

Table 1 contains the percentage of “detected” trees for each plot. The trees are subdivided into 3 layers with respect to h_{top} . The lower layer contains all trees below 50 % of h_{top} , the intermediate layer refers to all trees between 50 % and 80 % of h_{top} , and, finally, the upper layer contains the rest of the trees.

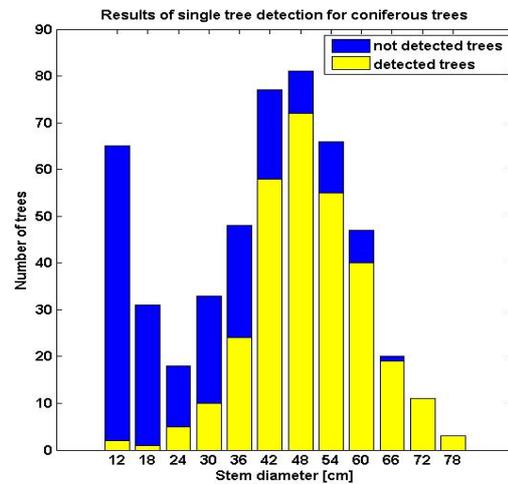


Figure 8. Number of detected coniferous trees in dependence on the DBH

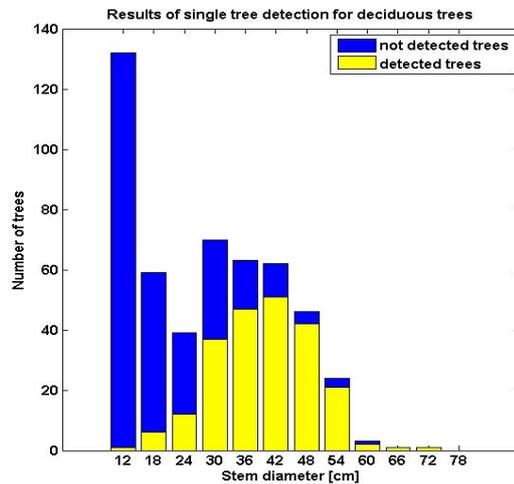


Figure 9. Number of detected deciduous trees in dependence on the DBH

First, we evaluate the detection rate of trees that are derived from the CHM without stem detection and hence refer to a local maximum in the CHM. Table 1 shows in general that most of the detected trees are in the upper layer. The detection rate varies between 56 % and 94 % while the overall detection rate for all plots amounts to 74 %. In comparison, in the intermediate layer and lower layer the detection rate is considerably smaller. Especially, in the lower layer only a few trees can be found since most of these trees are covered by

taller trees. The mean number of false detected trees amounts to 5 % and indicates a remarkable reliability. When applying the stem detection we get an overall improvement of the detection rate in the intermediate layer of 8 % and in the upper layer of 4 %. However, no improvement is achieved in the lower layer since (i) laser hits at the stem of small trees happen rarely, (ii) the base height h_{base} is inaccurate for trees beneath taller trees, and (iii) some trees have no clear base height since their green branches start close to the ground.

Moreover, the improvement of the detection rate is different in the individual plots. For instance, in the plots 55, 56, 58 and 81 (multi-layered stands), which contain more trees in the lower layer, no additional stems could be detected, whereas the plots 57 and 74 are significantly improved with more than 15 %. In general, the number of false detected trees gets slightly larger, most notably in the plot 93. Finally, the figures 8 and 9 show clearly that the detection rate of trees with a larger stem diameter at breast height (DBH) is considerably high for coniferous and deciduous trees. On average, the detection rate for coniferous trees is 61 % and for deciduous trees 44 %, respectively. Finally, table 2 shows the absolute positional improvement of the trees derived from the stem positions $(x_{stem}^{det}, y_{stem}^{det})$ and the position of the reference trees. As expected, the mean positioning error of deciduous trees gets better by 26 % which corresponds to 43 cm. The overall improvement of the tree position amounts to 24 cm which is equivalent to 21 %.

	Without "stem detection"	With "stem detection"
Mean positioning error coniferous	0.80 m	0.70 m
Mean positioning error deciduous	1.65 m	1.22 m
Total mean positioning error	1.16 m	0.92 m

Table 2. Accuracy of the tree position without and with stem detection

4. DISCUSSION

Conceptually, the presented approach to single tree detection from airborne LIDAR data goes one step further than existing methods by using the CHM and additional information inside the tree. It leads to an improvement of the detection rate of single trees in the intermediate and upper forest layer by detecting tree stems. This refinement of the detection rate could be expected since (i) in many cases neighbouring trees do not appear as two clear maximums in the raw data and (ii) the smoothing of the CHM blurs the maximums. Apparently, as already pointed out by some other authors (e.g. Solberg et al., 2006), the smoothing of the reconstructed CHM influences the quality of the single tree detection considerably. In general, a strong smoothing deteriorates the detection rate but decreases the number of false detected trees. However, a slight smoothing is necessary to avoid the finding of several wrong local maximums in one tree. Note that the surface reconstruction method in our approach works like an edge preserving filter which automatically adapts to local surface structures. However, a weak smoothing factor of 0.3 is still needed since otherwise the surface gets too noisy. The second advantage of the presented method is that the position of detected trees is significantly improved. This is also not very surprising since the

intersection of the detected tree stem with the DTM must be more precise than the tree position derived from the CHM maximum. Thirdly, the stem detection checks the hypothesis of stem positions which have been derived from the CHM. The restrictions of the approach are that only trees in the upper and intermediate forest layer can be additionally detected. It fails in the lower layer where stem hits are rare and stems points can not be clearly clustered. Also, the tree height still depends on the highest point found in the raw data contained in the stem cylinder V_{stem} . Thus, in all cases where the tree belonging to the detected stem is covered by a taller tree the derived tree height h_{tree} can be erroneous. Moreover, so far we have not implemented to go back to the raw data and to find a new segmentation of the tree crowns using the stem information.

Heurich (2006) reported recently on segmentation results using conventional first/last pulse data captured – like our data – in the Bavarian Forest National Park. He applied the segmentation approach developed by Persson et al. (2002). In detail, the detection rates for coniferous trees were 51%, for deciduous trees 40 %, and on average 45 %. These results are slightly worse than our results, but cannot be directly compared, because the test plots of that study were not exactly the same. However, the comparison is important since it refers to the same forest type. The study of Persson et al. (2002) reports on a detection rate of 71 % of all trees with a DBH larger than 5 cm for a Scandinavian forest dominated by spruce and pine. The positional accuracy of the stem positions was 51 cm. Recently, Solberg et al. (2006) presented a study for a structurally heterogeneous spruce forest with an overall detection rate of 66 % and commission rate (=false detections) of 26 % if the CHM is smoothed 3 times with a Gaussian filter of size 30 cm.

5. CONCLUSIONS

The study presents a novel single tree detection based on a combined surface reconstruction and stem detection. The results attained in heterogeneous forest types show clearly that the detection rate and position of single trees can be improved in the upper and intermediate layer. Future research should be focussed on (i) the improvement of the segmentation of the tree crowns using the stem information (ii) the analysis of the point's intensity and pulse width the waveform decomposition provides. More important, a rigorous 3D segmentation of the trees based on the points and the intensity and pulse width should be tackled to fully reconstruct the trees even in the lower layer.

6. LITERATURE

Brandtberg, T., 2007, Classifying individual tree species under leaf-off and leaf-on conditions using airborne lidar. *ISPRS Journal of Photogrammetry and Remote Sensing*, 61, pp. 325 – 340.

Heijden, F. van der, Duin, R.P.W., Ridder, D. de, Tax, D.M.J., 2004. *Classification, parameter estimation and state estimation – An engineering approach using MATLAB*. John Wiley & Sons Ltd, The Attriium, southern Gate, Chichester, West Sussex PO19 8SQ, England.

Heurich, M., 2006, Evaluierung und Entwicklung von Methoden zur automatisierten Erfassung von Waldstrukturen aus Daten flugzeuggetragener Fernerkundungssensoren. *Schriftenreihe des Wissenschaftszentrums Weihenstephan für*

Ernährung, Landnutzung und Umwelt der Technischen Universität München und der Bayerischen Landesanstalt für Wald und Forstwirtschaft, Forstlicher Forschungsbericht München, Nr. 202, ISBN 3-933506-33-6. <http://meadiatum2/ub.tum.de/> (Accessed February 18, 2007).

Holmgren, J., Persson, Å., 2004, Identifying species of individual trees using airborne laser scanner. *Remote Sensing of Environment* 90 (2004) 415-423.

Hyypä, J., Hyypä, H., Litkey, P., Yu, X., Haggren, H., Rönholm, P., Pyysalo, U., Pikänen, J., Maltamo, M., 2004, Algorithms and methods of airborne laser scanning for forest measurements. *Proceedings of the ISPRS working group VIII/2 Laser-Scanners for Forest and Landscape Assessment*, Volume XXXVI, PART 8/W2, 3 – 6th October, Freiburg, pp. 82 – 89.

Hyypä, J., Kelle, O., Lehtikoinen, 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*, 39:969-975.

Jutzi B., Stilla U. 2005. Waveform processing of laser pulses for reconstruction of surfaces in urban areas. In: Moeller M, Wentz E (eds) 3th International Symposium: Remote sensing and data fusion on urban areas, URBAN 2005. International Archives of Photogrammetry and Remote Sensing. Vol 36, Part 8 W27.

Krzystek, P., Wild, D., 1992, Experimental accuracy analysis of automatically measured digital terrain models. *Robust*

Computer Vision: Quality of vision algorithms. Förstner, Ruhwedel (ed.). Wichman Verlag, Karlsruhe, pp. 372 -390.

Persson, A., Holmgren, J. and Söderman, U., 2002. Detecting and measuring individual trees using an airborne laserscanner. *Photogrammetric Engineering & Remote Sensing* 68(9), pp. 925–932.

Pyysalo, U., Hyypä, H., 2002. Reconstructing Tree Crowns from Laser Scanner Data for Feature Extraction. In *ISPRS Commission III, Symposium 2002 September 9 - 13, 2002, Graz, Austria*, pages B-218 ff (4 pages).

Reitberger, J., Krzystek, P., Stilla, U. 2006. Analysis of Full Waveform Lidar Data for Tree Species Classification. *Symposium ISPRS Commission III "Photogrammetric Computer Vision" PCV06*, 20 – 22th September, Bonn, Germany.

Solberg, S., Naesset, E., Bollandsas, O. M., 2006. Single Tree Segmentation Using Airborne Laser Scanner Data in a Structurally Heterogeneous Spruce Forest. *Photogrammetric Engineering & Remote Sensing*, Vol. 72, No. 12, December 2006, pp. 1369-1378.

Vincent, L., Soille, P., 1991, Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, Vol. 13, No. 6, June 1991, pp. 583-598.

Plot	55	56	57	58	74	81	91	92	93	94	95	All plots	
Average age [years]	240	170	100	85	85	70	110	110	110	110	110		
Size [ha]	0.14	0.23	0.10	0.10	0.30	0.30	0.36	0.25	0.28	0.29	0.25	2.60	
Number of trees per ha	830	340	450	440	700	610	260	170	240	250	240	390	
Number of trees in lower layer	76	31	0	10	11	29	31	13	7	15	6	229	
Number of trees in intermediate layer	22	19	4	4	33	59	11	3	2	4	0	161	
Number of trees in upper layer	18	27	41	30	165	96	53	27	58	54	53	622	
Percentage of deciduous [%]	5	10	0	14	29	100	76	100	64	97	10	49	
Without "stem detection"	Detected trees lower layer [%]	1	7	0	0	0	7	8	0	0	0	3	
	Detected trees intermediate layer [%]	32	42	25	25	6	9	0	50	0	0	13	
	Detected trees upper layer [%]	56	74	81	73	69	85	85	72	94	89	74	
	Total number of detected trees [%]	16	39	76	52	55	30	51	56	64	70	80	49
	False detected trees [%]	1	1	0	0	1	8	8	19	9	7	3	5
With "stem detection"	Detected trees lower layer [%]	1	7	0	0	0	7	8	0	0	0	3	
	Detected trees intermediate layer [%]	32	42	75	25	30	9	33	100	0	0	21	
	Detected trees upper layer [%]	56	74	93	73	77	89	85	76	96	91	78	
	Total number of detected trees [%]	16	39	91	52	66	30	53	58	69	71	81	53
	False detected trees [%]	1	1	0	0	7	9	12	19	13	8	12	7

Table 1. Detection of trees in the reference plots

BUILDING CHANGE DETECTION FROM DIGITAL SURFACE MODELS AND MULTI-SPECTRAL IMAGES

Franz Rottensteiner

Cooperative Research Centre for Spatial Information, Dept. of Geomatics, University of Melbourne,
723 Swanston Street, Melbourne VIC 3010, Australia - franzr@unimelb.edu.au

Commission III, WG III/4

KEY WORDS: Change Detection, Digital Surface Models, Data Fusion

ABSTRACT:

A new method for building change detection from Digital Surface Models (DSM) and multi-spectral images is presented. The DSM can be generated from Airborne Laserscanner (ALS) data or by image matching techniques. From the multi-spectral image, the Normalised Difference Vegetation Index (NDVI) is computed and used in the change detection process. The workflow of the method consists of two stages. In the first stage, the DSM, the NDVI, and surface roughness parameters derived from the DSM are used in a classification technique based on the Dempster-Shafer theory for data fusion. In the case of ALS data, the height differences between DSMs created from first and last pulse data can also be considered. This technique is used to detect buildings. In the second stage of processing, these building detection results are compared to an existing building data base, and changes between the existing data base and the new data set are determined. This paper focuses on the second processing stage, in which the actual change detection is carried out. The method is designed to classify buildings and building parts as being *confirmed*, *changed*, *new*, or *demolished*. The change detection method considers the facts that the original data and the building detection results can have a different topology and that small differences between the data from the two epochs might be caused by different levels of generalisation or errors caused by a misalignment or insufficient resolution of the sensor data. Examples for the performance are given using DSMs generated both from ALS data and by image matching, highlighting the different properties of these data for building change detection.

1. INTRODUCTION

1.1 Motivation and Goals

Automatic building detection has been an important topic of research in Photogrammetry for more than a decade. In order to achieve results compliant with mapping scales of 1:1000-1:10000, multi-spectral aerial imagery and/or airborne laser scanner (ALS) data have been used (Matikainen et al., 2003; Rottensteiner et al., 2007). In many industrialised countries there exist 2D maps or building databases. Due to the dynamic nature of industrialised societies, building data collected at a certain time become outdated rather quickly. To speed up the production cycle for keeping such databases up-to-date, it is desirable to automate the detection of areas where buildings have changed. It is the goal of this paper to describe such a method for building change detection based on Digital Surface Models (DSMs) and multi-spectral imagery. This method is based on an adaptation of previous work on automatic building detection (Rottensteiner et al., 2007). The building detection results are compared to the existing map, and a classification is carried out to determine the changes between the two epochs. This classification has to take into account deviations between the two data sets that might be due to different degrees of generalisation and to small registration errors between the original data captured at the two epochs. It also has to consider the fact that the topology of the existing map and the automatically detected buildings might be different (Ragia and Winter, 2000). The method will be applied to DSMs generated from ALS data and by digital image matching, showing how well these different data sets are suited for the purpose of change detection for updating existing building data bases.

1.2 Related Work

There are two basic approaches to the problem of change detection (Vosselman et al., 2004). If original data are available for two different epochs t_1 and t_2 , change detection can be carried out by comparing the two datasets and inferring changes from the differences detected in the original data. This has been applied in the past to detect changes in buildings after earthquakes based on their different appearance in DSMs generated by ALS (Murakami et al., 1998). In the second scenario, a map or a digital data base is available for epoch t_1 and original data are only available for epoch t_2 . Change detection is carried out to keep the map up-to-date. In order to infer changes, the original map can be compared to the new sensor data directly. If ALS data are used for change detection, changes in buildings will result in height differences between a 3D city model and the new DSM. However, if the original data are only available in 2D, this cannot be directly exploited. This problem can be circumvented by detecting the objects of interest independently in the new sensor data and comparing the object detection results to the original map (Vosselman et al., 2004). This strategy does not take into consideration the fact that a building exists in the original map, it is not unlikely that there will still be a building at epoch t_2 .

Matikainen et al. (2003, 2004) detect buildings in ALS and aerial image data and then compare the results to building segments from an existing map. The method applies rule-based classification techniques separately to the building detection results and to the buildings of the existing map, based on the percentage of the area of a building that overlaps with any building in the other data set. Buildings in the existing data set are classified as *detected*, *partly detected*, and *not detected*,

whereas buildings in the new data set are classified as *new*, *enlarged*, or *old*. Small errors caused by alignment errors and by the generalisation of the existing map are considered by selecting thresholds different from 100% for classifying a building as *old* or *detected*. A common visualisation of these two classification results is presented on a per-pixel basis, but no further object-based analysis is carried out.

Vögtle and Steinle (2004) present a method for building change detection from ALS data that is based on a comparison of an existing map and the DSM corresponding to that map with the newly detected buildings and the corresponding DSM. Change detection results in a classification of buildings as *not-altered*, *new*, *demolished*, *added-on*, or *reduced*. In a first step, the overlap ratio between buildings is evaluated for both the old and the new data set. This leads to an initial classification of buildings as *new*, *demolished*, or *other*. The buildings not yet classified have a correspondence in the other data set, and are further classified as *not-altered*, *added-on*, or *reduced* based on an analysis of the height changes of the DSM in the areas covered by corresponding segments. Thus, this is an example for a change detection algorithm that uses the original data for both epochs. The advantage that this method can also detect height changes is contrasted by the problem of actually having access to such data.

Vosselman et al. (2004) present a method for comparing an existing map with the results of a building detection technique using ALS data. They give a list of errors that might result in differences between the existing map and the newly extracted buildings, namely generalization, random noise, systematic alignment errors, and object selection, the latter being a variety of generalization. Morphologic filters are applied before the comparison of the two data sets to compensate for errors caused by generalization, and an offset between corresponding segments is determined by a matching technique for coping with alignment errors.

In (Rottensteiner et al., 2007), a method for building detection by the fusion of ALS data and a normalised difference vegetation index (NDVI) derived from the red and the infrared bands of a multi-spectral image was presented. This method has been modified so that it can be applied to building change detection. The main focus of this paper will be on describing a new change detection method that is based on a comparison of the existing map and the results of building detection.

1.3 Method Overview

Building change detection requires a DSM generated either from ALS data or by image matching and, optionally, an NDVI image generated from a geocoded multi-spectral image. The DSM, the NDVI, and surface roughness parameters derived from the DSM are used in a building detection technique based on the Dempster-Shafer theory for data fusion. In the case of ALS data, the height differences between DSMs created from first and last pulse data can also be used. This classification technique has been modified so that it can also consider the existing building data base. An outline of the modified building detection method will be given in Section 2. In a second processing stage, the results of building detection are compared to an existing building data base and changes between the existing data base and the new data set are determined. This method can handle data sets of different topology and will be described in Section 3. Section 4 will present first results, whereas conclusions will be drawn in Section 5.

2. BUILDING DETECTION

The input to the method for building detection presented in (Rottensteiner et al., 2007) comprises up to four data sets generated from the raw data in a pre-processing stage. The minimum set of input data consists of a DSM grid and a Digital Terrain Model (DTM). The DSM can be derived from ALS data or by image matching. For the experiments described in Section 4, the DTM was derived from the DSM by hierarchic morphologic filtering (Rottensteiner et al., 2005). If ALS data are used, a DSM grid representing the height differences between the first and the last pulse can also be used. The fourth data set that can be used in building detection is the NDVI.

Building detection is based on the theory of Dempster-Shafer for data fusion. In Dempster-Shafer fusion, the output of a set of “sensors” is used for a classification process in which n classes $C_j \in \theta$ are to be discerned. Denoting the power set of θ by 2^θ , a *probability mass* fulfilling certain constraints has to be assigned to every class $A \in 2^\theta$ (i.e., also to any combination of the original classes) by each sensor. The probability masses from the individual sensors can be combined, and from these combined probability masses, two parameters can be computed for each class: the *Support* of a class is the sum of all masses assigned to that class, and the *Plausibility* sums up all probability masses not assigned to the complementary class of A . The accepted hypothesis $C_a \in \theta$ is determined according to a decision rule. The Dempster-Shafer theory also provides a measure for the *Conflict* in the sensor data (Klein, 1999).

Building detection starts with a Dempster-Shafer fusion process carried out for each pixel of the DSM to achieve a classification of the input data into one of four classes: buildings (B), trees (T), grass land (G), and bare soil (S), thus $\theta = \{B, T, G, S\}$. The model for the distribution of the evidence from each sensor to the four classes assumes that each sensor i can separate two complementary subsets of θ , U_{Ci} and U_{Ci}^c . The probability mass $P_i(x_i)$ assigned to U_{Ci} by the sensor i depending on the sensor output x_i is modelled to be equal to a constant P_l for $x_i < x_l$. For $x_i > x_u$, it is modelled to be equal to another constant P_u , with $0 \leq P_l < P_u \leq 1$. Between x_l and x_u , the probability mass is modelled by a cubic parabola with horizontal tangents at $x_i = x_l$ and $x_i = x_u$. The probability mass $[1 - P_i(x_i)]$ is assigned to U_{Ci}^c . No other assumptions about the distributions of the sensor data with respect to the classes are required. The combined probability masses are evaluated for each pixel, and the pixel is assigned to the class of maximum support. Originally, up to five “sensors” could be used in this process. The height differences ΔH between the DSM and the DTM help to distinguish elevated objects from the ground, thus $U_{CAH} = B \cup T$. Two surface roughness parameters computed from the first derivatives of the DSM, namely strength R and directedness D , are also used in the classification process. Large values of surface roughness are typical for trees, thus $U_{CR} = T$ and $U_{CD} = T$. However, D is only used if R differs significantly from 0. The height differences ΔH_{FL} between the first and the last pulse DSMs also distinguish trees from other classes: $U_{CAHFL} = T$. Finally, the NDVI is an indicator for vegetation, thus $U_{CNDVI} = T \cup G$. The uncertainty of the NDVI in shadow areas can be considered by modulating the probability masses depending on the standard deviation of the NDVI. In (Rottensteiner et al., 2007) it was shown how the parameters of the model for the probability masses can be selected. The classification results are improved by a post-classification technique aiming at re-classifying isolated pixels and pixels having a high conflict value. Initial building regions

are determined as connected components of “building pixels”. A second Dempster-Shafer fusion process is applied to these regions, using four cues representing average values for each building region to eliminate regions actually corresponding to trees. The result of building detection is the label image of the remaining building regions. The boundaries of these building regions can be extracted in vector format, too.

For the purpose of building change detection, a further optional input “sensor” was integrated into the first Dempster-Shafer fusion process. In many scenes, the amount of change will not be high, so that the original map gives an indication where buildings are to be expected. Thus, if a pixel is situated inside a building in the original map, it is more likely still to be inside a building at epoch t_2 than not. On the other hand, if the pixel is not inside a building in the original map, it is also more likely that it is not inside a building at epoch t_2 . Let the probability that a pixel inside a building in the original map is still a part of a building at epoch t_2 be denoted by P_B , and the probability that a pixel not being inside a building in the original map is not inside a building at epoch t_2 by P_{-B} . For a pixel inside a building in the original map, a probability mass equal to P_B can be assigned to class B , and $(1 - P_B)$ to class $T \cup G \cup S$. If the building is not inside a building in the original map, P_{-B} is assigned to class $T \cup G \cup S$ and $(1 - P_{-B})$ is assigned to class B . Thus, a bias is introduced by the original map, which can especially help to confirm small buildings that might otherwise be classified as trees. In the current implementation, $P_{-B} = P_B$ is assumed, and the user has to specify the value for P_B . Typically, P_B is chosen between 60 % and 75 %. The user can also decide not to consider the original map in the classification if the amount of change in the scene is high.

3. BUILDING CHANGE DETECTION

Change detection is based on a comparison of label images: the “existing label image” generated from the existing map and containing the labels $l^e \in L^e$, and the “new label image” generated by the building detection method and containing the labels $l^n \in L^n$. The goal of change detection is three-fold:

1. A classification of the buildings in the existing map as either *confirmed*, *changed*, or *demolished*;
2. The detection of buildings that are found to be *new*;
3. A delineation of the outlines of both the *changed* and the *new* buildings, showing *demolished* and *new* building parts for the *changed* buildings.

Since the two data sets will usually have different topologies, these goals are achieved in a procedure consisting of two stages. The first stage is a topological clarification of the new data set by matching its labels to those of the existing label image in order to achieve topological consistency between the two epochs. This is followed by the actual classification to detect the changes. The output of change detection consists of a “change map” showing the actual changes between the epochs according to the classification results described above, and a label image describing the state at epoch t_2 .

3.1 Topological Clarification

For each co-occurrence of two labels $l^e \in L^e$ and $l^n \in L^n$, the overlap ratios $p_{ne} = n_{n \cap e} / n_n$ and $p_{en} = n_{n \cap e} / n_e$ are computed, where $n_{n \cap e}$ is the number of common pixels assigned to the

region l^n in the new label image and to l^e in the existing label image, n_n is the total number of pixels assigned to the region l^n in the new label image, and n_e is the total number of pixels assigned to l^e in the existing label image. First, marginal correspondences, i.e. correspondences that do neither contribute significantly to l^n nor to l^e , are eliminated. Using a user-defined threshold t_m (typically, 10 %), all correspondences with $p_{ne} < t_m$ and $p_{en} < t_m$ are eliminated. As a result, a set of correspondences between labels $l^e \in L^e$ and $l^n \in L^n$ is obtained. Figure 1 shows two sets of labels L^e and L^n , with correspondences depicted by lines. If the topology of the two data sets were identical except for *new* or *demolished* buildings, each label would have no or one line connecting it to a label in the other data set. As the topology is not identical, six cases must be distinguished (cf. Figure 1):

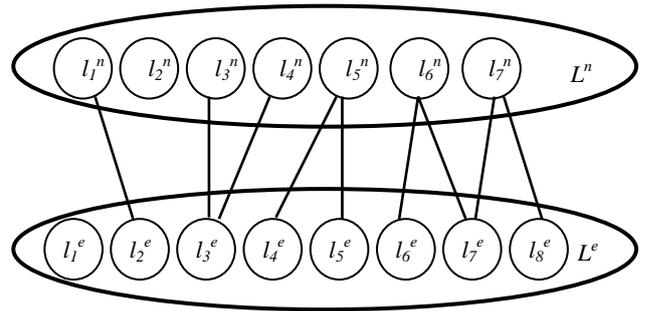


Figure 1. Two sets of labels L^e and L^n and correspondences between labels $l_i^e \in L^e$ and $l_j^n \in L^n$.

1. A label in L^e not having any corresponding label in L^n indicates a *demolished* building (e.g. l_1^e in Figure 1).
2. A label in L^n not having any corresponding label in L^e indicates a *new* building (e.g. l_2^n).
3. A label in L^e having exactly one corresponding label in L^n (e.g. l_2^e and l_1^n) indicates a *confirmed* or a *changed* building.
4. If a label in L^e has M corresponding labels in L^n (e.g. l_3^e , l_3^n and l_4^n), the original building is split into several labels. It can be either *confirmed* or *changed*. The splitting can be caused by the actual demolition of building parts, or it can just be caused by height discontinuities within a building.
5. If a label in L^n has N corresponding labels in L^e (e.g. l_5^n , l_4^e and l_5^e), several existing buildings are merged. Again, this might be the result of some actual changes or not. It typically occurs with terraced houses having identical roof heights.
6. If a set of M labels $l_i^e \in L^e$ corresponds to a set of N labels $l_j^n \in L^n$ (e.g. l_6^e , l_7^e , l_8^e , l_6^n and l_7^n), buildings or building parts are both split and merged, and there are ambiguities with respect to the correct correspondence of some of the new labels (Figure 2).

In all except the first two cases, there can additionally be building parts that overlap with the background in the other label image, i.e., *new* or *demolished* parts of *changed* buildings.

In order to obtain a classification on a per-building level for all existing buildings, the ambiguous case 6 has to be resolved first. Assuming the topology of the existing label image to be correct, the new label image has to be changed so that no ambiguities occur. This can be achieved by splitting any label l_j^n that corresponds to more than one label in L^e and/or has a significant overlap with the background. This shall be explained using the example in Figure 2. The existing label image, shown in the upper row, contains two labels l_1^e and l_2^e . The second row shows the results of building detection. There are altogether five

labels l_j^n , one of which (l_2^n) overlaps with both l_1^e and l_2^e . The label l_4^n also contains a new building that has been constructed in a gap within the building l_2^e . First, new buildings or building parts are identified. For this purpose, a binary image of those building pixels in the new label image that correspond to the background in the existing label image is generated. This binary image highlights the *new* building pixels, but it is also affected by noise at the building outlines. A morphological opening filter is used to remove the noise. The size of the structural element is the minimum size of a change that can be detected, and is typically chosen to correspond to one to three times the sensor resolution. If there remain "white" pixels in the filtered image, new labels corresponding to new building parts can be detected by a connected component analysis. A label image L^c combining the existing labels L^e and these new labels is created. Each label in L^n corresponding to more than one label in L^c is split so that each of the new labels corresponds to exactly one label of L^c . In order to compensate for smoothing effects of the morphological filter at the fringes of new building parts, the Voronoi diagram of L^c is used to assign pixels to one of the new labels. The third line of Figure 2 shows the results of this splitting process. Two new labels were added to L^n . The label l_6^n corresponds to a new building part, and the label l_7^n is the result of splitting off the part of l_2^n that corresponds to l_2^e .

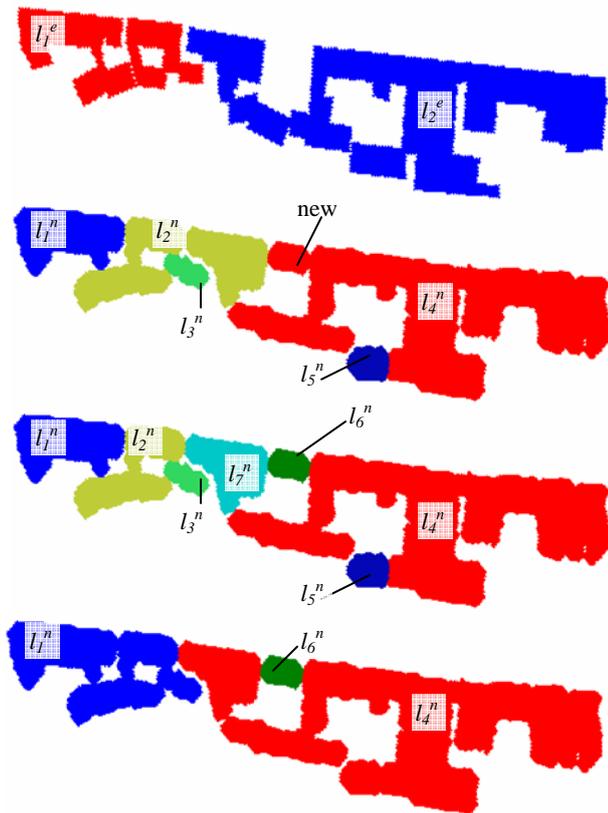


Figure 2. First row: Existing map. Second row: Results of automatic building detection. Third row: New label image after resolving ambiguities. Fourth row: Results of topological clarification.

Having resolved all ambiguities, the case corresponding to merged labels is resolved. Labels can be merged because the buildings are close to each other (left part of Figure 3), or because a new building has been constructed between them (right part of Figure 3). In a similar process as described above, new building labels are detected, and the merged label in L^n is

split into several new labels, each corresponding either to a new building or to exactly one label in the existing map.

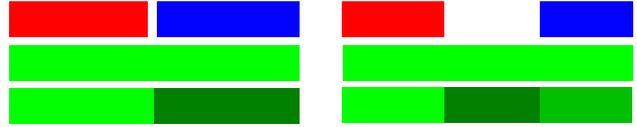


Figure 3. Clarifying the case of merged labels. Upper row: Existing map with two buildings. Second row: Results of building detection. Third row: Clarified label image. Left: the merged buildings are close to each other. Right: a new building was detected.

Finally, the case corresponding to split labels is analysed. We want to merge all new labels corresponding to an existing label if the fact that these labels are separate in the new label image is not the result of a larger building part having been demolished. For that purpose, the new labels are grown by morphologic closing (i.e., the binary image of building pixels is closed morphologically, and each building pixel in the closed image is assigned to the label found in the Voronoi diagram of the new label image). If two labels are found to be neighbours in the closed label image, the two labels are merged. If this is not the case, the separation is assumed to have been caused by the demolition of a building part, and the original labels are maintained. The last row in Figure 2 shows the results of topological clarification after merging of the split labels. As a result of topological clarification, an improved version $L^{n_{imp}}$ of the new building label image L^n is obtained, with some of the original labels in L^n having been split and others having been merged. Each of the labels in $L^{n_{imp}}$ corresponds either to exactly one label in L^e or to none (Figure 4). Each of the labels of L^e corresponds to one or more labels of the improved version of $L^{n_{imp}}$ or to none. There remaining split cases (one label of L^e corresponding to more than one label in the improved version of $L^{n_{imp}}$) all correspond to *changed* buildings.

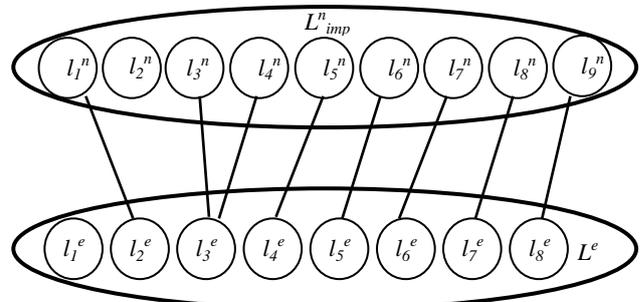


Figure 4. Two sets of labels L^e and $L^{n_{imp}}$ and correspondences between labels $l_i^e \in L^e$ and $l_j^n \in L^{n_{imp}}$ after topological clarification.

3.2 Classification of Changes

After topological clarification, the actual change detection is carried out. Again, the percentages of overlap are computed for each co-occurrence of two labels $l^e \in L^e$ and $l^n \in L^{n_{imp}}$. Marginal correspondences are eliminated, which results in a correspondence graph like the one shown in Figure 4. A building is classified as *new* if its label $l^n \in L^{n_{imp}}$ does not have a correspondence in L^e . An existing building is classified as *demolished* if its label $l^e \in L^e$ does not have a correspondence in $L^{n_{imp}}$. For all remaining building labels $l^e \in L^e$, two binary images are created: a binary image of *demolished* pixels (i.e., pixels assigned to l^e in the existing map, but not to any of the

corresponding labels in L_{imp}^n), and a binary image of *new* pixels (i.e., pixels assigned to any of the labels corresponding to l^c in L_{imp}^n , but not assigned to l^c in the existing map). Both are filtered by morphological opening. The size of the structural element is chosen to be identical to the one used in topological clarification. If neither *demolished* nor *new* pixels remain after morphological opening, the existing building is classified as *confirmed*; otherwise, it is classified as *changed*. For *changed* buildings, connected components in the binary images of *demolished* and *new* pixels are considered to correspond to *demolished building parts* and *new building parts*. Finally, two images representing the change detection results are generated:

1. A *change map* using different colours for *new*, *demolished*, *confirmed*, buildings as well as for *confirmed*, *new*, or *demolished parts* of *changed* buildings.
2. A label image representing the new state.

For generating the label image representing the new state, there are two options. The first option is to use the improved version of the new label image. As an alternative, the original building outlines can be used for *confirmed* buildings, whereas for *changed* buildings, the outlines can be a combination of the original outlines for those building parts that have not been changed and the *new building parts*. The second option is to be preferred if the original map is more accurate, e.g. if it is a cadastral map generated by a geodetic survey.

4. EXPERIMENTS

The method described in this paper was tested using two data sets. The first data set, captured over Fairfield (NSW), consisted of ALS points with a nominal spacing of 1.2 m. The first and the last laser pulses as well as the intensity of the returned signal were recorded. In addition, a colour orthophoto was available. From the red band of that orthophoto and the intensity of the signal, a “pseudo-NDVI” image could be generated. DSMs of a grid width $\Delta = 1$ m were generated for both the first and the last pulse data. For a part of the Fairfield data set, the outlines of the buildings were determined by photogrammetric plotting with a planimetric accuracy of ± 0.2 m. In order to simulate actual changes, the label image generated from these outlines was modified by both adding and removing buildings or building parts. This modified building map was used as the existing map in change detection. The size of the test area was about 500×400 m². The second data set was provided by EuroSDR. It consisted of a DSM generated by image matching, a colour orthophoto and an orthophoto representing the infrared band, and an existing building data base in the form of a binary building image. Both the DSM and the orthophoto had a resolution of $\Delta = 0.501$ m. The size of the test area was about 1100×1100 m². The DSM was very noisy, especially in the shadow regions, where the outlines of buildings were smoothed. The numerical resolution of the DSM heights was identical to the planimetric resolution Δ . Thus, the only height values occurring were full multiples of Δ . Along with the uncertainties of the DSM in shadow areas, this was the reason why surface roughness was of no use for building detection.

Building detection using the method outlined in Section 2 was applied to both data sets. In Fairfield, standard parameters described in (Rottensteiner et al., 2007) were used for the height differences between DSM and DTM, the height differences between first and last pulse DSMs, the NDVI, and the two

surface roughness parameters. The existing map was also considered in the classification process, using $P_B = P_{-B} = 75$ %. In Toulouse, we used the DSM, the NDVI, the directedness of surface roughness (but not the strength), and the existing map, using $P_B = P_{-B} = 60$ % and selecting the other parameters in the way described in (Rottensteiner et al., 2007). As described in Section 3, the results of building detection were used to detect changes between the original map and the new data. The resulting change maps, generated at the resolution of the respective DSM grids, are presented in Figure 5.



Figure 5. Change maps for Fairfield (above) and Toulouse (below). Ochre/yellow: *confirmed* buildings/building parts. Blue/light grey: *demolished* buildings/building parts. Red/green: *new* buildings/building parts.

In Fairfield, changes affecting the main buildings are detected correctly. All *new* buildings (red areas in Figure 5) and building parts (green) in the data set were detected. All *new* buildings are correct, and so are the majority of the *new* building parts. The few incorrect *new* building parts are the result of an over-estimation of the building extents. Of course, the algorithm cannot really discern whether a larger area found to be new is a new building or only a new building part. In case of doubt, a new building is assumed. All *demolished* buildings (blue) and building parts (light grey) were detected. All except two *demolished* building parts are correct. However, with *demolished* buildings, the trend observed in (Rottensteiner et

al., 2007) is confirmed: For the main buildings, the change detection results are correct, but for smaller structures, the results become more and more uncertain. Thus, the small buildings classified as *demolished* in the back yards, mostly garden sheds and garages, are actually too small to be detected by the method and hence have to be considered as false negatives. A comparison of the label image representing the new state to the reference label image was carried out to derive quality parameters on a per-pixel level. *Completeness*, *correctness*, and *quality* (Rottensteiner et al., 2005) were determined for the building pixels of the label image representing the new state. *Completeness* was 95.0 %, thus 5 % of the building pixels in the reference data set were not detected; these are mostly small buildings in the back yards. *Correctness* was 97.9 %, thus only 2.1 % of the detected building pixels were incorrect. The overall *quality* was 93.1 %. These numbers represent the actual classification accuracy and are not affected by misalignment of the data sets, because the original outlines were used for *confirmed* buildings.

In Toulouse, the results are not as good as for Fairfield. Again, it can be observed that small buildings are not detected in the new data set. Many major changes are detected correctly, e.g. the groups of *new* and *demolished* buildings in the south-west of the scene. There are two very large areas of false detections: the large green area in the northeast corner of the scene merges a correctly detected new building with a parking lot, and the large red area in the east is actually a sports field. In both cases, as well as in the case of some building parts erroneously classified as *demolished*, the poor quality of the DSM in these areas (height variations larger than 3 m in essentially horizontal areas) contributed to these false classifications. Note also the obvious over-estimation of the large building complex in the western part of the scene in the shadow areas (i.e., north of the building). Other problems were related to trees partly overhanging buildings and to the general lack of ground points in the forested areas, the latter causing errors in the DTM generation process. On a per-pixel basis, *completeness* was determined to be only 89.3 %, and *correctness* was even poorer (68.1 %). Without the two problematic areas, *correctness* was still only 76.5 %. These data are partly afflicted by errors in the reference data. They were determined by digitization in the orthophoto and, thus, are not as accurate as the Fairfield reference data. In any case, a comparison of the results for the two data sets shows the importance of using a high-quality DSM in the classification process.

5. CONCLUSIONS

A new method for building change detection was presented. It combines DSMs, DSM roughness parameters, an NDVI image, and an existing map in a classification process based on Dempster-Shafer fusion. Comparing the building detection results to the existing map, a topological clarification is carried out, and the changes between the existing map and the results of building detection are classified. The classification takes into account the fact that small differences are likely to be caused by errors of building detection at the building outlines. The results of change detection are presented so that the user can easily assess which buildings are *confirmed*, *new*, *demolished*, or *changed*, and in case of *changed* buildings also the nature and extent of these changes. Given the importance of the DSM in the classification process, it is not astonishing that the method works considerably better with DSMs derived from ALS data compared to DSMs generated by image matching. However,

this may be partly the result of the specific matching algorithm used to generate the Toulouse DSM. In the future, the method will be tested further using the remaining EuroSDR data sets. Further work will concentrate on improving the geometrical quality of the building outlines by image edges, because the building outlines are much better defined in the image data than in a DSM, especially if the latter was generated by matching.

ACKNOWLEDGEMENTS

The Fairfield data set was provided by AAMHatch, Australia (www.aamhatch.com.au). The Toulouse data set was one of three that were distributed by EuroSDR in their test on the detection of unregistered buildings for updating cadastral databases (buildingsdetection.free.fr).

REFERENCES

- Klein, L., 1999. *Sensor and data fusion, Concepts and applications*. 2nd edition, SPIE Optical Engineering Press.
- Matikainen, L., Hyypä, J., Hyypä, H., 2003. Automatic detection of buildings from laser scanner data for map updating. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Dresden, Germany, Vol. XXXIV, Part 3/W13, pp. 218-224.
- Matikainen, L., Hyypä, J., Kaartinen, H., 2004. Automatic detection of changes from laser scanner and aerial image data for updating building maps. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Istanbul, Turkey, Vol. XXXV, Part B2, pp. 434-439.
- Murakami, H., Nagakawa, K., Shibata, T., Iwanami, E., 1998. Potential of an airborne laserscanner system for change detection of urban features and orthoimage development. In: *The International Archives of the Photogrammetry and Remote Sensing*, Stuttgart, Germany, Vol. XXXII, Part 4, pp. 422-427.
- Ragia, L., Winter, S., 2000. Contributions to a quality description of areal objects in spatial data sets. *ISPRS Journal for Photogrammetry and Remote Sensing* 55 (3), pp. 201-213.
- Rottensteiner, F., Trinder, J., Clode, S., Kubik, K., 2005. Using the Dempster-Shafer method for the fusion of LIDAR data and multi-spectral images for building detection. *Information Fusion* 6 (4), pp. 283-300.
- Rottensteiner, F., Trinder, J., Clode, S., Kubik, K., 2007. Building detection by fusion of airborne laserscanner data and multi-spectral images: Performance evaluation and sensitivity analysis. *ISPRS Journal for Photogrammetry and Remote Sensing*, 62(2), pp. 135-149.
- Vögtle, T., Steinle, E., 2004. Detection and recognition of changes in building geometry derived from multitemporal laserscanning data. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Istanbul, Turkey, Vol. XXXV, Part B2, pp. 428-433.
- Vosselman, G., Gorte, B., Sithole, G., 2004. Change detection for updating medium scale maps using laser altimetry. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Istanbul, Turkey, Vol. XXXV, Part B3, pp. 207-212.

AUTOMATIC VELOCITY ESTIMATION OF TARGETS IN DYNAMIC STEREO

Norbert Scherer, Michael Kirchhof, Rolf Schäfer

FGAN-FOM Research Institute for Optronics and Pattern Recognition, Gutleuthausstr. 1, 76275 Ettlingen, Germany

KEY WORDS: camera calibration, homography, parameter estimation, position estimation, velocity estimation, structure from motion

ABSTRACT:

In this paper we focus on a stereo system consisting of one stationary and one moving platform. The stationary platform detects motions and gives cues to the moving platform. When an object is in the field of view of both sensors the object's position, velocity, and acceleration are estimated from stereo. To prove the results a second stationary sensor platform is available to compute a reference from traditional stereo. We describe investigations about the accuracy and reliability concerning the automatic velocity estimation of targets in image sequences of multi-ocular systems with emphasis on moving sensor platforms. In order to make the assessment more comprehensible, we discuss a stereo-system consisting of one stationary and one moving platform. Therefore in this context dynamic stereo means that the stereo basis and relative orientation are varying within the image sequence.

We present image sequences originating from different sensors at different positions including one moving sensor that were recorded during the same time span. Since all-day capability is required for warning systems predominantly infrared (IR) sensors were employed. We took image sequences of scenarios with one or more vehicles. The vehicles moved with typical velocities in natural surrounding. Their distances to the sensors was in the close range. The trajectories of all sensors and vehicles were registered with GPS-recorders to obtain ground-truth data.

1 INTRODUCTION

Modern automatic surveillance and warning systems need to ensure the safety of high value targets. To achieve results that guarantee adequate protection against different attacks high quality classifications of threats are essential. Incorporating the movement patterns of active objects appearing in the field of view of the surveillance system into the analysis of potential threats improves the reliability of the classification results. This is true even more, if the information about the movement of the objects is of high quality. The required high quality information could be obtained by a reconstruction of the trajectory of the objects in time and space, which can be done with multi-ocular stereo vision. Furthermore, a more precise definition of the threat is obtained, if the accuracy of the reconstructed trajectory allows the derivation of the three-dimensional velocity and acceleration of each object. Especially for objects moving directly towards the sensor, the three-dimensional position, velocity or even acceleration of the object are of highest interest, since they result in more robust features than shape, size, texture or intensity for the analysis of the threat. Since also moving objects need protection, e.g., convoys of vehicles, the necessity to discuss a moving sensor platform arises. In order to make the assessment more comprehensible, we discuss a stereo-system consisting of one stationary and one moving platform.

The analysis starts with detection and tracking of objects in each individual image sequence. Additionally the trajectory of the moving sensor is estimated by Structure-From-Motion methods (Hartley and Zisserman, 2000), (Kirchhof and Stilla, 2006). The accuracy of the trajectory and relative orientation is improved by sliding bundle adjustment over up to 100 subsequent frames. The resulting trajectory is then matched with the corresponding GPS-track of the vehicle carrying the moving sensor. This ensures that all cameras can now be described in the same coordinate system. Afterwards the correspondence problem is solved and the three-dimensional trajectories and velocities of the observed objects are reconstructed (Scherer and Gabler, 2004). The accuracy of the reconstructed trajectories in space and time is assessed by

comparison to the recorded GPS-data. Furthermore to analyze possible performance degradations arising from the movement of sensors, we compare the results of the moving stereo system with the results of a stereo system consisting of two stationary sensors. The solution to the correspondence problem and the trajectory and velocity estimation of the observed object is identical for the stationary and the dynamic case.

1.1 Related Work

Typical approaches for the estimation of the trajectory and velocity of moving objects assume that the objects move on an observed plane. In this case a monocular image sequence is sufficient to determine the trajectory and velocity of a moving object. One typical example may be (Reinartz et al., 2006). Reinartz et al. compute a georeferenced image sequence which allows to compute the position and velocity directly from image displacements. Nistér presented quite different work about dynamic stereo (Nistér et al., 2004) introducing a system of a stereo camera mounted on a vehicle. The advantage of this system is that the stereo basis and relative orientation remain constant over the sequence although the sensors are moving. The approach presented here is based on (Scherer and Gabler, 2004) where the range and velocity of objects at long ranges were computed from triangulation. Caused by the long range application the contribution focuses on discretization effects in the range and velocity estimation. For such applications the relative rotations between the sensors is very important while the relative positioning error can be very large without affecting the results. Therefore positioning the sensors with GPS was sufficient in that work.

Our application is in close range where the relative positioning error induces very large disparities. Therefore the registration of the stationary sensor positions was improved by adjustments over many GPS-measurements supported by distance measurements. Additionally the registration of the orientation was done by the comparison of the sensor's view with the view of a virtual camera computed from laser measurements taken from a helicopter.

1.2 Notation

We describe experiments in which we employed different vehicles (\mathcal{V}_1 , \mathcal{V}_2 and \mathcal{V}_3). The vehicles \mathcal{V}_1 and \mathcal{V}_2 were monitored with three different cameras (\mathcal{C}_1 , \mathcal{C}_2 and \mathcal{C}_V). Cameras \mathcal{C}_1 and \mathcal{C}_2 were stationary, whereas camera \mathcal{C}_V was mounted on top of vehicle \mathcal{V}_3 , which followed \mathcal{V}_1 and \mathcal{V}_2 . \mathcal{V}_1 was driving in front of \mathcal{V}_2 . The positions of the vehicles were tracked with Global Positioning Systems (GPS). The GPS-receivers are described by the letter \mathcal{G} . The connection to the vehicle, which's position is measured, is established by the use of an index that corresponds to the vehicle, i. e. the GPS-system in the vehicle \mathcal{V}_1 is depicted by \mathcal{G}_1 . To obtain further ground-truth information we used the information of a GPS-system \mathcal{G}_H mounted on a helicopter \mathcal{V}_H , which performed laser measurements of the area in which our experiments took place.

2 DATA ACQUISITION

The data described in this paper had been recorded during a measurement campaign realized by the authors and other members of FGAN-FOM. The campaign took place at the end of September 2006.

Three infrared cameras (\mathcal{C}_1 , \mathcal{C}_2 and \mathcal{C}_V) were used as imaging sensors. The cameras \mathcal{C}_1 and \mathcal{C}_2 were from AIM INFRAROT-MODULE. Camera \mathcal{C}_V was from FLIR Systems. The technical details of the cameras are summarized in table 1.

Ground truth data were recorded using the Global Positioning System (GPS). The vehicles \mathcal{V}_1 and \mathcal{V}_2 employed portable GPS-systems GPSMAP 76CSx (\mathcal{G}_1 and \mathcal{G}_2). \mathcal{V}_3 , which carried the Camera \mathcal{C}_V , was equipped with a GPS-mouse system (\mathcal{G}_3). The portable GPS-systems were also used to determine the positions of the stationary cameras \mathcal{C}_1 and \mathcal{C}_2 and some additional outstanding points in the terrain.

Furthermore the terrain was scanned with a Riegl LMS Q650 laser scanner, which was mounted on a Helicopter (\mathcal{V}_H) of type Bell UH-1D. The scans produced several overlapping stripes containing height profiles of the terrain. \mathcal{V}_H was equipped with an Inertial Measurement Unit (IMU) and a GPS-antenna (\mathcal{G}_H). The signals from both sensors were processed by a special computer in such a way that position, orientation and acceleration of the helicopter are known during the data acquisition phase. Further details of the helicopter equipment can be found in (Hebel et al., 2006).

3 TRIANGULATION AND VELOCITY ESTIMATION

We are now going to describe our general approach. First we show the general procedure of obtaining a three-dimensional track for the case of two stationary cameras. Second the necessary modifications to expand the approach to the case of moving sensors are depicted.

Generally the approach is divided into two steps. In the first step the image sequences of each sensor are processed separately. The

Camera	FOV	Spectrum	No. of Pixels
\mathcal{C}_1	$17.5^\circ \times 13.3^\circ$	4.4 - 5.2 μm	640×480
\mathcal{C}_2	$18.7^\circ \times 14.1^\circ$	2.0 - 5.3 μm	384×288
\mathcal{C}_V	$20.0^\circ \times 15.0^\circ$	3.0 - 5.0 μm	320×256

Table 1: Technical Data of the used IR-cameras.

results of this step are then used as input to the second step. The second step combines the results of the analysis of the two image sequences and constitutes the desired three-dimensional track of the object of interest.

3.1 Stationary Case

The first step of creating a three-dimensional track applies an Infrared Search and Track (IRST) algorithm to each of the image sequences. This algorithm starts with pre-processing the images to correct for sensor specific inhomogeneities. Afterwards the image sequences are integrated to increase the signal-to-noise-ratio. In the resulting sequences point like objects are tracked, so that two-dimensional tracks of these objects are created in each image sequence.

Figures 1, 2 and 3 show examples of the 'point-like objects' as seen from the cameras \mathcal{C}_1 , \mathcal{C}_2 and \mathcal{C}_V . The images have been taken at the same time. The point-like objects found by the application of the IRST-algorithm are marked with rectangles. Please notice that not all of the marked points in one image must have a corresponding mark in any of the other two images. On the other hand the blue rectangle in each image marks a point that has correspondences in the other images. That point belongs to the back of vehicle \mathcal{V}_1 . An example of a two-dimensional track resulting from one object is given in figure 4.

The second step uses the two-dimensional tracks that have been created by the IRST-algorithm and reconstructs the three-dimensional trajectories of objects by combining corresponding two-dimensional tracks from the image sequences. For this reconstruction the knowledge of the camera's position and orientation are important. Further details and a theoretical discussion of the accuracy and reliability of this approach can be found in (Scherer and Gabler, 2004).

3.2 Dynamic Case

In the dynamic case one camera is moving during the observation. Therefore the second step of the analysis procedure is modified in such a way that the possible changes of the positional information (position and orientation) of the camera are considered. These information are obtained with Structure from Motion methods as described later in 4.3.

Since these methods only return relative positional informations, they have to be transformed into our reference frame by an Euclidean transformation. This is done by fitting the whole track of the moving camera \mathcal{C}_V to the whole track of the sensor carrying vehicle \mathcal{V}_3 obtained from the GPS-system \mathcal{G}_3 .

Due to the variance of the positional information obtained by our GPS-receivers a final translation of the whole track of the moving camera is needed. This translation is obtained by comparing the position of the moving camera at one point in time with the position where it is seen in the corresponding image of one of the stationary cameras.

4 CALIBRATION AND SYNCHRONIZATION

The measurement of the velocity of vehicles requires knowledge about the time at which the vehicle is at a certain point. Since we want to estimate the velocity from different cameras and compare the results with ground-truth-data the data-streams of all sensors need to be 'synchronized' not only in time but also in space.



Figure 1: Image from camera C_1 . In the center of the picture vehicle V_1 can be seen. Vehicle V_2 follows V_1 . The rectangles mark 'objects' for which two-dimensional tracks had been created by the IRST-algorithm.

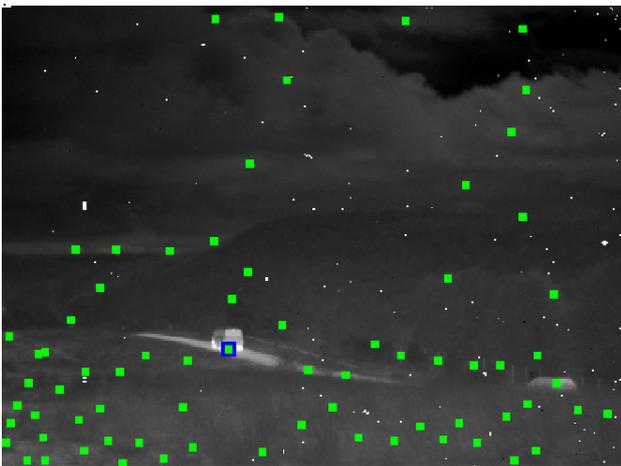


Figure 2: Image from Camera C_2 taken at the same time as the image shown in figure 1. Rectangles mark 'objects' that had been found by the application of the IRST-algorithm.

4.1 Spatial Registration

The positions of the stationary cameras C_1 and C_2 were established by combing all available position measurements (GPS, position information derived from the laser scans of the helicopter and some additional distance measurements), stating a minimization problem for the position and distance differences between the measurements and solving it with the Levenberg-Marquardt algorithm.

As a result of this procedure we obtained the camera positions within a precision less than half a meter, which is much better than the variance of one of our single GPS-measurements.

Now that the position of the stationary cameras had been fixed we obtained the orientation of the cameras with the virtual overlay technique. By this we use the data of the height profiles from the laser scanner to produce 'images' of a camera with a virtual reality tool. These pictures are then compared with the real camera image. The comparison is done in an overlay of the real and the virtual image. The parameters of the virtual camera are then manually modified until a reasonable conformance between both images is reached. An example of an overlay image is seen in figure 5.



Figure 3: Picture from camera C_V taken at the same time as the picture 1. 'objects' resulting from the processing of this video stream by the IRST-algorithm are again marked as rectangles.



Figure 4: Same image as in figure 1. Here a two-dimensional track, as it resulted from the application of the IRST-algorithm to the whole image-sequence of camera C_V , is overlaid. The track belongs to the 'object' that had been marked with a blue rectangle in the center of figure 1.

4.2 Temporal Registration

For the temporal registration, we need to synchronize our cameras to a global time, e.g. GPS-time. Fortunately in order to achieve the synchronization we only need to determine one constant time-shift for each camera, since each image sequence is equipped with timestamps of a local clock. We identified this constant at that parts of the image-sequences that show starting vehicles, since this incident could be identified with high precision in the GPS-time-stream.

4.3 Structure from Motion

Since the GPS-data contains no information about the orientation (rotation) of the sensor an image-based reconstruction approach is required. The reconstruction is initially computed independently from the available GPS-data and is improved by sliding bundle adjustment which takes the GPS-data into account. We assume that the internal camera parameters are known for example by the use of self-calibration techniques like (Zhang, 1999).

In the first step points of interest are detected in the first frame (Förstner and Gülch, 1987). These points are then tracked

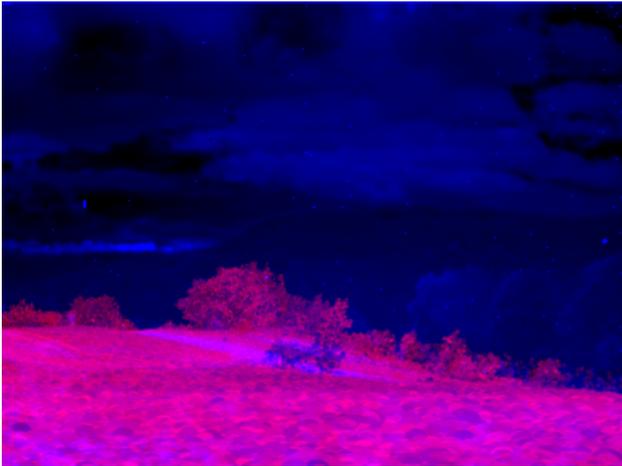


Figure 5: Example of an overlay picture used to obtain the orientation of camera C_2 . Reddish scene parts belong to the image as seen from a virtual camera with data based on the laser measurements. Bluish parts belong to the IR-image from the camera C_2 .

through the sequence based on template matching (Lucas and Kanade, 1981). Outliers of the tracking process are detected with RANSAC (random sample consensus) (Fischler and Bolles, 1981) for homographies based on our previous work (Kirchhof and Stilla, 2006). Every tenth frame the tracked point set is improved by applying the point of interest detection to the current frame.

The relative orientation can now be computed from the essential matrix using the five point algorithm of Nist'er (Nist'er, 2003). This relative orientation enforces the triangulation of the corresponding 3d-point set. Subsequent frames can now be stitched to the reconstruction by linear regression followed by non-linear minimization of the reprojection error using Levenberg-Marquardt (McGlone et al., 2004). The 3d-point set can now frequently be updated in a robust way by retriangulation again using RANSAC.

As mentioned above we refine the reconstruction with bundle adjustment over the latest one hundred frames using Levenberg-Marquardt taking the GPS-data into account. Although we used a tracking and matching strategy the computed tracks may be corrupted by slightly moving objects or drifts of the tracker. The Huber robust cost function (Huber, 1981) reduces the influence of such errors while it is still convex. Therefore no additional local minima are induced by it.

5 EXPERIMENTS

For the comparison of the stationary case with the dynamic case we tracked vehicle \mathcal{V}_1 with all three cameras. The figures 1 to 3 show images of the sequences. Within these images the blue rectangles mark the objects in the images that we used to reconstruct the three-dimensional trajectory of \mathcal{V}_1 . For the camera C_1 the two-dimensional track of the object marked with a blue rectangle is shown in figure 4.

5.1 Stationary Case

The result of the evaluation of the stationary cameras C_1 and C_2 is visible in figure 6 as a blue line. Because of the field of view of the camera C_2 only the last part of the track is visible. In that range the position of the track coincides very well with the real track of the vehicle.

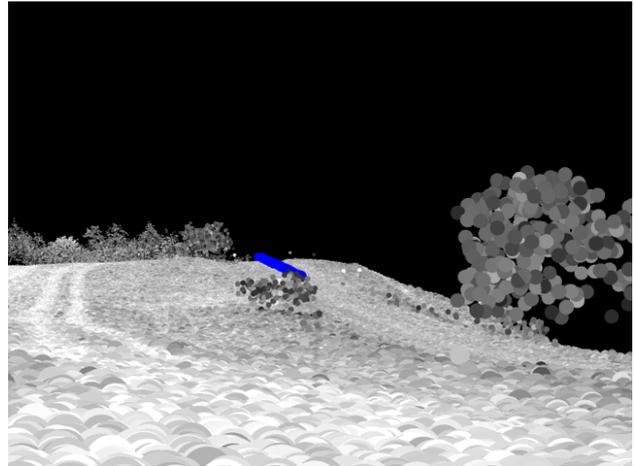


Figure 6: Reconstruction of the three-dimensional trajectory of vehicle \mathcal{V}_1 (blue line) based on the images from the stationary cameras C_1 and C_2 .

5.2 Dynamic Case

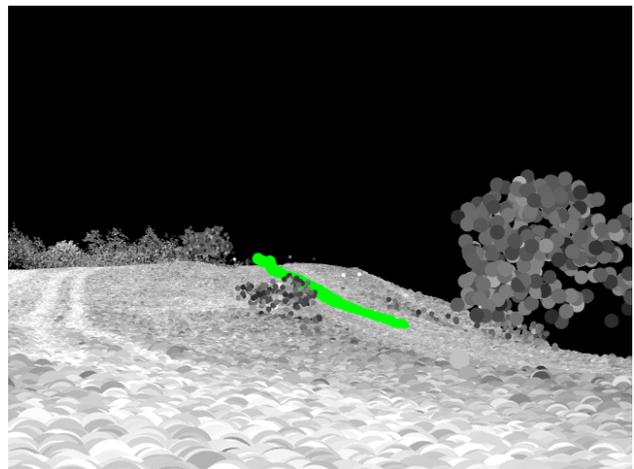


Figure 7: Three-dimensional trajectory of \mathcal{V}_1 (green line) reconstructed from the cameras C_1 and C_V .

For the case including one moving camera C_V , which was the main purpose of our investigation, the resulting three dimensional track of vehicle \mathcal{V}_1 is shown in figure 7 as a green line. Again the track is in good accordance with the track of vehicle \mathcal{V}_1 , as could be seen by comparison with the two-dimensional track shown in figure 4.

5.3 Comparison

A more quantitative comparison of the reconstruction results with the real trajectory of \mathcal{V}_1 as the one presented in the previous sections 5.1 and 5.2 is shown in figure 8 as a top view. The center of the depicted coordinate system coincides with the position of the camera C_1 . The vehicle \mathcal{V}_1 moves from the upper left corner to the lower right corner. It is obvious that in the stationary (blue dots) and the dynamic case (green dots) the reconstructed positions match well with the GPS-measurements (black dots). At the end the dynamic track shows problems arising from instabilities in computation of the position and orientation of the moving camera C_V . These problems result in reconstructed positions showing a backward moving vehicle in contradiction to the real trajectory of \mathcal{V}_1 . Possibly the problem arises from the same source as the fact that the pitch of the camera C_V needed a correction of about 1° before the data could be processed. Up to that

point the results are quite good, as seen in the transition to the track reconstructed from the stationary cameras.

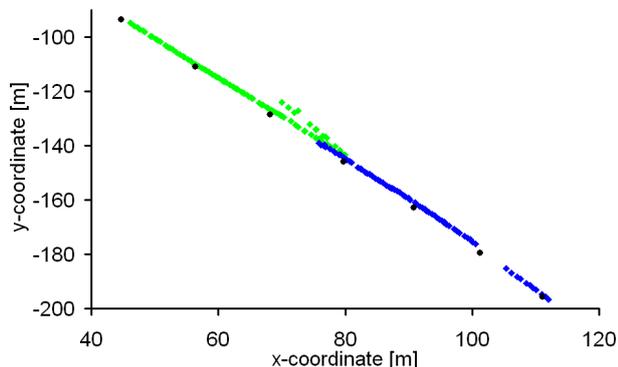


Figure 8: Comparison between the positions obtained in the stationary (marked by green dots) and the dynamic (blue dots) case with GPS-measurements. Black dots represent the positions of vehicle \mathcal{V}_1 as measured with \mathcal{G}_1 .

Based on the good three-dimensional reconstruction of the position it is now possible to derive the velocity of the vehicle \mathcal{V}_1 . The velocity is calculated as a running linear regression over 15 time points. Figure 9 shows the results as green dots for the dynamic case and blue dots for the stationary case. The mean velocity value from the GPS-data is shown as a black line. For both cases the velocities vary around the mean value obtained from the GPS-data. Obviously the velocity values from the dynamic case are distorted by the reconstruction problems of the trajectory mentioned in the previous paragraph, which start at 26.5 seconds in figure 9. On the other hand it is seen, that without these problems the stationary data are a good continuation of dynamic ones. Furthermore the figure shows that the variation of the velocity obtained from the dynamic case (up to 26.5 seconds) is equal to the variation of the velocity obtained from the stationary case.

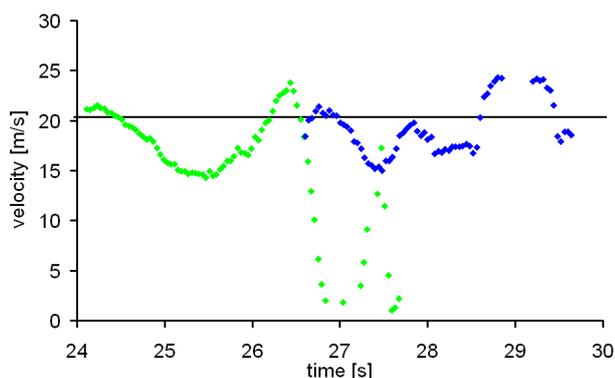


Figure 9: Comparison of the velocities obtained from the stationary (marked by green dots) and the dynamic (blue dots) case. The horizontal black line represents the mean velocity obtained from the GPS-data.

6 CONCLUSION

We described the results obtained from systems for automatic velocity estimation from stereo data. One of the systems had a fixed stereo basis, the other had a varying stereo basis since the second camera of the stereo pair was mounted on a moving vehicle. It has been shown that the described approach is applicable in principle, provided that a high quality registration of all necessary data is available.

The structure from motion methods need to be supported by additional metrical information, e. g., GPS in our case. But the obtainable results seem to depend strongly on the quality of this additional information. In our case the velocity calculation was quite good in the beginning, but failed when the position estimates obtained by the structure from motion approach break down. Further detailed investigations will be necessary to find the cause of this failure.

7 OUTLOOK

As we pointed out above the structure from motion approach is the bottleneck of the presented work. The 3d registration can be improved by considering not only the relative orientation of the monocular moving sensor but also the relative orientation between the moving and the stationary sensor. This is in general a wide base stereo approach. Therefore the used descriptors for points of interest have to be replaced by rotational and scale (and in the optimal case affine) invariant descriptors like SIFT (Lowe, 2004) - or MSR - features. Additional improvement can be obtained by detecting the moving objects and exclude them from further processing.

8 ACKNOWLEDGMENTS

The work described above has been supported by BMVg (Bundesverteidigungsministerium, German Ministry of Defense) and BWB (Bundesamt für Wehrtechnik und Beschaffung, German Office for Armament and Procurement). The assistance of the WTD 81 (Wehrtechnische Dienststelle für Informationstechnologie und Elektronik) for preparation and realization of the measurement campaign at Greiding in September 2006 is gratefully acknowledged.

REFERENCES

- Fischler, M. A. and Bolles, R. C., 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the Association for Computing Machinery* 24(6), pp. 381–395.
- Förstner, W. and Gülch, E., 1987. A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres of Circular Features. In: *ISPRS Intercommission Workshop, Interlaken*.
- Hartley, R. and Zisserman, A., 2000. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, UK.
- Hebel, M., Bers, K.-H. and Jäger, K., 2006. Imaging sensor fusion and enhanced vision for helicopter landing operations. In: J. J. G. Jacques G Verly (ed.), *Signal Processing, Sensor Fusion and Target Recognition XV*, Vol. 6226, SPIE.
- Huber, P. J., 1981. *Robust Statistics*. John Wiley Publishers.
- Kirchhof, M. and Stilla, U., 2006. Detection of moving objects in airborne thermal videos. *ISPRS Journal of Photogrammetry and Remote Sensing* 61, pp. 187 – 196.
- Lowe, D., 2004. Distinctive image features from scale-invariant keypoints. *Int. J. of Computer Vision* 60(2), pp. 91–110.
- Lucas, B. T. and Kanade, T., 1981. An Iterative Image Registration Technique with an Application to Stereo Vision. In: *Proc. of Image Understanding Workshop*, pp. 212–130.

McGlone, J. C., Mikhail, E. M. and Bethel, J. (eds), 2004. Manual of Photogrammetry. 5th edn, American Society of Photogrammetry and Remote Sensing.

Nistér, D., 2003. An efficient solution to the five-point relative pose problem. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '03) 02, pp. 195.

Nistér, D., Naroditsky, O. and Bergen, J., 2004. Visual odometry. Vol. 1, pp. I-652-I-659.

Reinartz, P., Lachaise, M., Schmeer, E., Krauß, T. and Runge, H., 2006. Traffic monitoring with serial images from airborne cameras. ISPRS Journal of Photogrammetry and Remote Sensing 61, pp. 149 – 158.

Scherer, N. and Gabler, R., 2004. Range and velocity estimation of objects at long ranges using multiocular video sequences. International Archives of Photogrammetry and Remote Sensing 35(Part B5), pp. 741–746.

Zhang, Z., 1999. Flexible Camera Calibration by Viewing a Plane from Unknown Orientations. In: International Conference on Computer Vision (ICCV'99), Corfu, Greece, pp. 666–673.

TERRA SAR-X VALUE ADDED IMAGE PRODUCTS

Nadine Schmidt^a, Marco Weber^a, Johannes Raggam^b, Karlheinz Gutjahr^b, Andreas Wimmer^b

^a Infoterra GmbH, Claude-Dornier-Strasse, 88090 Immenstaad, Germany Nadine.Schmidt@infoterra-global.com,

^b Institute of Digital Image Processing, Joanneum Research Forschungsgesellschaft mbH, Steyrergasse 17, 8010 Graz, Austria
johann.raggam@joanneum.at, karlheinz.gutjahr@joanneum.at, andreas.wimmer@joanneum.at

KEYWORDS: TerraSAR-X, value adding, spaceborne radar satellite, X-band, high-resolution imagery, enhanced image products, geo-information products, radar remote sensing

ABSTRACT:

The space mission TerraSAR-X is the first German space project implemented under a Public Private Partnership (PPP). Cooperation partners are the German Aerospace Centre (DLR) and EADS Astrium GmbH. Within this construct, DLR will be responsible for the scientific use of the TerraSAR-X data, whereas commercial marketing will be undertaken exclusively by Infoterra GmbH, a wholly-owned EADS Astrium subsidiary.

In a co-operation between Infoterra GmbH and Joanneum Research, Value Added products and processors have been developed for TerraSAR-X data. These products are mainly oriented at the area of interest or are mapping products which represent a higher level of image processing in terms of radiometric correction and orthorectification, mosaics, subsets and merges. In this paper, these products are described. Further, an insight into the automated and semi-automated production chain is provided.

1 INTRODUCTION

Commercial users of remote sensing data require detailed data adapted to their individual requirements, available quickly and reliably, independent of daylight and weather conditions. The design and performance of TerraSAR-X will precisely meet these requirements.

The Synthetic Aperture Radar (SAR) instrument of the spacecraft supplies detailed high-resolution radar imagery, day and night, under all weather conditions. The acquired data is the basis for a wide variety of products and services, such as sophisticated client-specific image interpretation, topographic maps up to a scale of 1:10,000, geo-spatial databases and terrain analysis used for a wide scope of applications. Other application sectors include environmental planning, land cover mapping, natural resource exploration, regional and urban development, crisis response and relief, insurance and risk assessment, as well as applications in border control, security, intelligence and defence.

Based upon TerraSAR-X data, a variety of different image products are offered to the customers. The TS-X image products are differentiated into two major product groups:

- **Basic Image Products** (or LIB Products), which are distributed by Infoterra GmbH for the commercial clients and by DLR for the scientific users.
- **Value Added Products** (VA Products) which are further distinguished in Enhanced Image Products and Geo-Information Products.

TS-X Value Added products and processors were defined and developed in a co-operation between Infoterra GmbH and Joanneum Research. The products are distributed by Infoterra GmbH only.

2 TERRA SAR-X BASIC IMAGE PRODUCTS

Basic Image Products are **satellite path oriented** data sets. They correspond to the CEOS Level 1b quality. The SAR raw data are processed to basic image products by the TerraSAR-X Multi Mode SAR Processor (TMSP), which has been developed by DLR. For each individual order, the customer can specify processing options, which determine

- the geometric and radiometric resolution,
- the geometric projection, and
- the auxiliary information and annotation.

All TerraSAR-X Basic Image Products are processed and delivered in radar brightness β_0 , compatible to the detected ground range products from ERS-1 and 2, ENVISAT/ASAR and RADARSAT. In contrast to ENVISAT and ERS, the complex slant range products are delivered in radar brightness as well.

With respect to the geometric projection and data representation of the data, the TS-X Basic Image Products are differentiated into four product types as follows:

- **Single Look Slant Range Complex (SSC)**, with amplitude and phase information in slant range geometry;
- **Multilook Ground Range Detected (MGD)**, is corrected from slant range to ground range projection;
- **Geocoded Ellipsoid Corrected (GEC)**, corrected to Universal Transversal Mercator (UTM) and Uniform Polar Stereographic (UPS) projection using the WGS84 ellipsoid with an average terrain height;
- **Enhanced Ellipsoid Corrected (EEC)**, corrected to UTM or UPS projection with a digital elevation model (DEM), e.g. the SRTM elevation models, which are available with a mesh width of 1 and 3 arcseconds, respectively.

These products can be delivered as either a radiometrically or a spatially enhanced product. The spatial extent corresponds to the area accessible by the TerraSAR-X standard scene. For further information on the TS-X Imaging Modes and the Basic Image Products see [1].

3 TERRA SAR-X VALUE ADDED PRODUCTS

The TS-X Value Added Products are mainly **oriented at the area of interest (defined by the customer) or are map products** which represent a higher level of image processing. These products are called **Enhanced Image (EI) Products** and will be available immediately after the satellite's commissioning phase. **Geo-Information (GI) Products**, which are based on more sophisticated processing and / or analysis methods for e.g. subsidence or change detection mapping will follow.

The Value Added Products are based on TerraSAR-X Basic Image Products and can be generated from new acquisitions, catalogue orders, or a combination of both. Due to a higher processing level, the products provide more processing options than the Basic Image Products and additional auxiliary information. With these products, the customer is not limited to satellite specific images and image strips but can order a full geometrically and radiometrically adjusted coverage of the selected region of interest. Further, it is possible to generate VA Products from existing VA Products, e.g. a Mosaic based on high level orthorectified images.

In the following section, the different Enhanced Image Products are described, followed by the main processing options.

3.1 Enhanced Image Products

3.1.1 ORI^{SAR} - Orthorectified image

The Orthorectified Image product is a highly accurate geocoded image including terrain correction. All terrain distortions inherent in satellite imagery, particularly in areas with a high relief, are expected to be removed. Due to the major influence of the Digital Elevation Model (DEM) on the pixel location accuracy of an orthorectified image, only high precision DEMs are used for ORI^{SAR} production. The result shows an increased pixel location accuracy in comparison to an EEC product.

The high precision DEMs used for ORI^{SAR} production can either be provided by the customer or be purchased from Infoterra GmbH. The ORI^{SAR} product is represented in map geometry. The standard map projections are UTM or UPS with WGS84 ellipsoid.

To achieve and validate high output quality for the various TS-X processing modes and resolution ranges, the following processing options can be utilized:

- optimization and validation of the underlying sensor models being used for ortho-rectification;
- automated GCP acquisition based on matching of control point chips, which are already available in a GCP chip data base;
- interactive GCP (Ground Control Points) acquisition using reference maps in order to increase the pixel location accuracy if necessary;
- dedicated quality assessment of the ORI^{SAR} output product.

The ORI^{SAR} is available with the radiometric representation in radar brightness β_0 like the Basic Image Products by default; further, an additional radiometric calibration (σ_0) or radiometric normalisation (γ_0) can be ordered (description of the radiometric corrections see RaN^{SAR}).

This product is expected to provide a higher level of geometric correction in comparison to Basic Image Products due to higher quality DEMs being used and sensor model optimization procedures being applied. It can be interpreted very quickly and is ideal for combination with other sources of information.

3.1.2 RaN^{SAR} - Radiometrically corrected image

The Basic Image Products GEC and EEC are delivered as radar brightness (β_0). The customer can order an additional radiometric calibration or radiometric normalisation of these products. The resulting product is called RaN^{SAR} (radiometrically corrected or normalized GEC or EEC image).

Both corrections compensate topography-induced effects such as local pixel scattering area and local incidence angle. Radiometric calibration and normalization will lead to a calculation of the sigma naught (σ_0) or gamma naught (γ_0) coefficient, respectively, depending on the customer's selection. Like the Basic Image Products GEC and EEC, the RaN^{SAR} is represented in map geometry. The standard map projections are UTM or UPS with WGS84 ellipsoid.

Radiometric correction may be preferred if the images are to be used for classifications which do not take angular dependencies of the SAR data into account. Further, it is also important in order to minimize the differences in the radiometry of the various images if several images of the same area or neighbouring areas are to be composed.

3.1.3 MC^{SAR} - Mosaic

To cover a geographical area larger than a standard scene, neighbouring geocoded or orthorectified images are seamlessly combined into one image. Possible input products are ORI^{SAR}, EEC or GEC, depending on the customer's localization accuracy requirements.

For mosaicing, various processing options can be utilized, some being optional, others mandatory:

- geometric quality check and optional geometric data adaptation in case of an insufficient geometric consistency. This can be done by means of low order polynomial transformation as well as matching-based image co-registration;
- radiometric quality check and optional adaptation of the input image data. Therefore, selected (critical) areas, e.g. large water bodies, can be excluded;
- automatic detection of specific mosaicing boundaries (cutlines), e.g. along natural edges or line structures like tree lines, streets or field borders. These cutlines are used to generate the mosaicing source mask, which provides information on the source input pixel to become the respective output pixel in the mosaic;
- use of source mask in mosaic production in order to avoid the visibility of cutting edges in the final image product. This will create a seamless mosaic without breaks or visible cutlines;
- dedicated quality assessment of the MC^{SAR} output product.

The MC^{SAR} is represented in map geometry. The standard map projections are UTM or UPS with WGS84 ellipsoid. The product provides seamless image information over a large area. It is quickly interpretable and combinable with other sources of information. Thus, it can be used for map sheet generation.

3.1.4 OI^{SAR} - Oriented Image

The Oriented Image is a subset of an orthorectified or geocoded image scene, mosaic or ascending / descending merge. The subset region is defined by the customer through an area of interest polygon or corner coordinates of the desired region. The product can either be characterized by the user defined area of interest or by a map sheet orientation according to relevant mapping standards or customer defined extensions. The product

is represented in map geometry, with the standard map projections UTM or UPS with WGS84 ellipsoid.

The OI^{SAR} can be combined with other sources of information, so it can directly be used e.g. for map sheet generation. The customer receives an image that covers the specified area of interest rather than satellite-typical image strips that may only cover the area in fractions.

3.1.5 ADM^{SAR} - Ascending/Descending Merge

Typical characteristics of SAR images acquired over rugged terrain are the radar layover, foreshortening and shadow effects, which are not useful in an ortho-rectified product. A reduction of these effects can be achieved by a combination of ortho-rectified SAR images which have been acquired from ascending and descending right looking orbits. Respective merging options are included in the VA processing suite, resulting in a so-called ADM^{SAR} product.

During ascending/descending merge, an ascending and a descending ortho image are combined in order to replace no-information areas – like layover and shadow areas – by information available in the respective other image. Furthermore, also for the information areas an “optimized” output pixel can be generated, based on decision and merge criteria, which consider e.g. the local incidence angle or pixel resolution relationship.

The scope of delivery includes the individual source images for the respective ascending and descending orbits in ortho-rectified geometric quality. Thus, EEC, ORI^{SAR} or MC^{SAR} generated from these products can be used as input to the ADM^{SAR} generation. The Source Image Mask (SOU) will present the origin of each pixel in the product. The ADM^{SAR} is represented in map geometry. The standard map projections are UTM or UPS with WGS84 ellipsoid.

The ADM^{SAR} is of particular interest for areas with steep mountain terrain, where shadow and layover can significantly disturb the analysis.

3.2 Envisaged Geo-Information Products

The following Geo-Information Products are currently planned and will be developed during the commissioning phase.

3.2.1 OM^{SAR} – OrthoMap

The OrthoMap consists of an orthorectified image with accurate image geometry within a map frame annotated by a legend. It is generated from Oriented Images (OI^{SAR}) which in turn have been produced from orthorectified images (EEC, ORI^{SAR} or MC^{SAR} products). The layout ranges from simple map frames to highly integrated cartographic products according to relevant mapping standards or customer specifications

3.2.2 CDM^{SAR} – Change Detection Map

CDM^{SAR} focuses on long-term monitoring of urban areas and settlements e.g. to detect surface sealing and urban sprawl. Such information can be used e.g. for urban monitoring and urban planning applications or map updates.

3.2.3 SUB^{SAR} – Subsidence Map

The subsidence map provides information on long-term surface displacement in urban areas and settlements. Such vertical surface displacement may be caused e.g. by tectonics, subsurface mining, or earthquakes, and the resulting maps can be used for risk diagnostics.

3.3 Processing options for Value Added Products

3.3.1 Radiometric correction

The TS-X Basic Image Products are delivered in radar brightness (β_0) by default. However, it is possible to order a radiometric correction for the GEC and EEC products – called RaN^{SAR} . For the ORI^{SAR} is also possible to generate radiometric corrected images. If one of these products is selected as input to ADM^{SAR} or OI^{SAR} , the respective output can also have a radiometric correction. The standard MC^{SAR} is delivered in radar brightness (β_0). It is possible to choose between two types of radiometric correction:

- **Radiometric calibration:** resulting in sigma naught (σ_0). The correction factor $\sin\alpha$ is applied to average pixel intensity determined from beta naught.
- **Radiometric normalization:** resulting in gamma naught (γ_0). The correction function $1/\cos\alpha$ is applied to sigma naught.

3.3.2 Auxiliary raster products

All TS-X Value Added Products are accompanied by auxiliary raster products, which can contain conversion, positional information and other ancillary information. Depending on the Value Added Product type, different auxiliary raster products are available.

Geocoded Incidence Angle Mask (GIM): The GIM contains information on the local incidence angle and on the location of radar shadowing and layover. This mask has the same coding and file structure as the GIM available for the EEC products [1]. The mask can be used for further processing, e.g. radiometric calibration using the incidence angle information. The GIM is available for ORI^{SAR} , RAN^{SAR} , and all other VA products derived from these products or EEC.

Incidence Angle Mask (IAM): The IAM is an alternative representation of the local incidence angle information. The values are not coded, thus the mask can be directly used for further processing such as radiometric calibration using the incidence angle information. The IAM is available for ORI^{SAR} , RAN^{SAR} , and all other VA products derived from them.

Layover and Shadow Mask (LSM): The LSM is the second part of the alternative representation of the GIM. It contains information on radar shadow and layover regions in the image. Like the IAM it can directly be combined with the image data. The LSM is available for ORI^{SAR} , RAN^{SAR} , and all other VA products derived from them.

Local Resolution Mask (RES): The RES identifies the actual ground resolution of the SAR system for each pixel resulting from the local topography and incidence angle. The RES is available for the ORI^{SAR} and products derived from it.

Source mask (SOU): The source mask gives a numeric value that allows the identification of the source input image for each output pixel. The Source Mask is optionally available for the MC^{SAR} , ADM^{SAR} and the OI^{SAR} .

Enumeration files: Enumeration files consists of the **Along Track** Enumeration file (**ALT**) and **Across Track** Enumeration file (**ACT**). Both files provide the original location in SAR (range-azimuth) geometry for each output pixel. These files are useful for conversions from slant range to geocoded geometries (and vice-versa), e.g. for geocoding additional products co-registered with the input image. The enumeration files are only available for ORI^{SAR} if it is produced with a DEM provided by the customer.

The following table gives an overview on the availability of Auxiliary Raster Products for TS-X Value Added Products.

Product	Auxiliary files						
	GIM	IAM	LSM	SOU	RES	ALT	ACT
ORI ^{SAR}	✓	✓	✓		✓	✓*	✓*
RAN ^{SAR}	✓	✓	✓				
MC ^{SAR}	✓*	✓*	✓*	✓	✓*		
OI ^{SAR}	✓*	✓*	✓*	✓*	✓*		
ADM ^{SAR}	✓*	✓*	✓*	✓	✓*		

* optional, depending on input products type (see description)

3.3.3 Projections

The standard cartographic projections for TS-X Value Added Products are Universal Transversal Mercator (UTM) and Uniform Polar Stereographic (UPS) with WGS84 ellipsoid. A transfer of the products to other projections is possible, depending on the required projection and customer order.

3.3.4 File Formats

The TS-X Value Added Products standard delivery format is GeoTiff. The annotation information is provided in xml format. COTS software (like Leica Geosystems Erdas Imagine, PCI Geomatics, etc) supports this format. However, it is also possible to order additional reformatting into e.g. ERDAS IMG, NITF, PCI PIX format.

3.3.5 Scaling

All TS-X Value Added Products are produced in 16bit. Sometimes the data may be reasonably reduced to 8bit, e.g. for visualization purposes for mapping applications, or if this scaling does not lead to significant loss of information detail. The customer can order an additional 8bit scaling of the VA Products image data; an automated or an interactive 8bit scaling option can be selected.

REFERENCES

- [1] M. Eineder and T. Fritz (coord.), "TerraSAR-X Ground Segment, SAR Basic Product Specification Document (TX-GS-DD-3302), Release 1.4, 06. October 2006
- [2] Infoterra GmbH, TSXX Value Added Product Specification (TSXX-ITD-SPE-0009, in preparation, 2007

IMPROVEMENT OF INTERIOR AND EXTERIOR ORIENTATION OF THE THREE LINE CAMERA HRSC WITH A SIMULTANEOUS ADJUSTMENT

Michael Spiegel

Photogrammetry and Remote Sensing, Technische Universität, München, 80333 Muenchen, Germany
spiegel@bv.tum.de, www.remotesensing-tum.de

WG I/2, WG III/2, WG III/4, WG III/5, and WG IV/3

KEY WORDS: Three line camera, interior orientation, exterior orientation, HRSC, combined adjustment

ABSTRACT

Since January 2004 the High Resolution Stereo Camera (HRSC) on board the ESA Mission Mars Express (MEX) is imaging the surface of planet Mars in color and stereoscopically in high resolution. As part of the entire data processing concept the Institute of Photogrammetry and GeoInformation (IPI) of the Leibniz Universität Hannover and the Department Photogrammetry and Remote Sensing (FPF) of the Technische Universität München are jointly responsible for the photogrammetric adjustment of the MEX-HRSC orientation data. This will be accomplished by registration of the HRSC data to the Mars Observer Laser Altimeter (MOLA) data of Mars Global Surveyor Mission using single strips or neighboring strips forming a block. With the result of the processing chain, high quality products such as Digital Terrain Models (DTMs), ortho-image mosaics and shaded reliefs can be derived from the imagery. In this paper the concept of simultaneous adjustment of interior and exterior orientation and results of photogrammetric point determination is described.

1 INTRODUCTION

The ESA mission Mars Express Mission (MEX) with the High Resolution Stereo Camera (HRSC) on board started the orbiting phase in January 2004. During the first three years more than 1400 stereo images were acquired.

The primary goal of the Photogrammetry and Remote Sensing (FPF) at the Technical University of Munich is to determine the exterior and interior orientation of HRSC orbiting planet Mars during Mars Express mission. In general, the classical photogrammetric point determination requires image coordinates of tie points, constant interior orientation, and ground control points (GCP). In case of HRSC on Mars Express tie points will be measured automatically in the images by means of image matching. Interior orientation is known from calibration. Observations for the exterior orientation will be derived from star observation, Inertial Measurement Unit (IMU) measurements, and doppler measurements. Unfortunately, these observations for the parameters of the exterior orientation will probably not be precise enough for a consistent photogrammetric point determination on a global level.

Additional control information is necessary in order to fit photogrammetrically derived object points into the existing reference system on Mars. On Mars there are only few precisely known points which can serve as classical GCPs. But there is a large number of ground points measured by MOLA. The characteristics of the laser points are, that they can not be identified in the images in an easy way. I.e., image coordinates of most of these points can not be measured, and therefore, it is not possible to treat them as normal GCPs in a bundle adjustment. As a remedy it is proposed to use control surfaces derived from the MOLA points.

2 DATA SOURCES

2.1 High Resolution Stereo Camera (HRSC)

The HRSC is a multi-sensor pushbroom camera consisting of nine Charge Coupled Device (CCD) line sensors for simultaneous high resolution stereo, multispectral, and multi-phase imaging. It has one panchromatic nadir channel, four panchromatic stereo channels, and four channels for color. The convergence angles between the nadir- and the stereo sensors are 21 and 14 gon. Figure 1 shows a HRSC Nadir image with tie points.



Figure 1: HRSC Nadir image with tie points(red)

The sensor arrays with 5184 active pixels each are arranged perpendicular to the direction of flight in one focal plane. The images are generated by catenating the continuously acquired line-images. The result is one image per sensor and orbit. One image strip includes all images of one orbit. The pixel size on ground of 12 m will be reached at an altitude of 270 km at pericentre and increase to 50 m at an altitude of 1000 km (Neukum and Hoffmann, 2000). At pericentre one image strip covers an area of about 60 km across trajectory. In general, the strip has a length of about 400 km up to 4000 km.

The three-dimensional position of the spacecraft is determined by the European Space Agency (ESA) applying a combination of doppler shift measurements (up to 15 times per orbit), acquisition of ranging data, triangulation measurements, and orbit analysis. The orbit accuracy at the pericentre is given as an interval of

maximum and minimum accuracy for the whole mission duration (Hechler and Yáñez, 2000). In direction of flight, perpendicular to the direction of flight, and radial direction the accuracy of the orbit is assumed to be about 1000 m.

The attitude of the spacecraft is commanded by European Space Operation Center (ESOC). With this predicted attitude, the orbiter adjust themselves using measurements of star tracker cameras and from an Inertial Measurement Unit (IMU). But, for photogrammetric point determination only the predicted attitude is available. Therefore, the accuracy of the nadir pointing results from a combination of attitude errors and navigation errors. The values for accuracies are 25 mdeg for all three rotation angles φ (pitch), ω (roll), and κ (yaw) (Astrium, 2001).

2.2 Mars Observer Laser Altimeter (MOLA)

In February 1999 the Mars Global Surveyor (MGS) spacecraft entered the mapping orbit at Mars. During the recording time (February 1999 to June 2001) the MOLA instrument acquired more than 640 million observations by measuring the distances between the orbiter and the surface of Mars. In combination with orbit and attitude information these altimeter measurements have been processed to coordinates of object points on the ground. Therefore, each orbit results in one track of MOLA points (Smith et al., 2001).

In addition to the surface described by the original, irregularly spaced MOLA track points NASA (Neumann et al., 2003) distributed a grid-based global Digital Terrain Model (DTM) which is derived from these MOLA points (see Figure 2). The accuracy of DTM is 200 m in planimetry and 10 m in height.

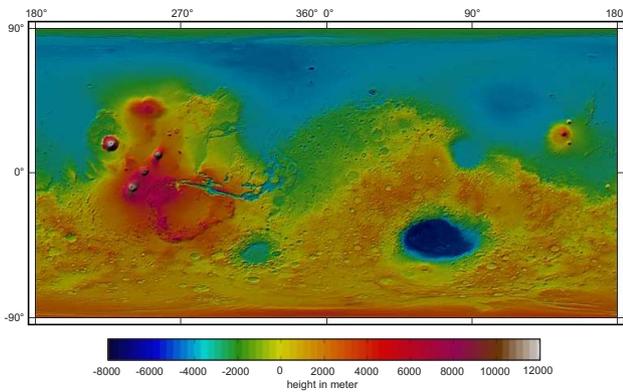


Figure 2: MOLA DTM

3 CONCEPT

In this Section the concept of automated measurement of image coordinates of tie points and the bundle adjustment will be described.

3.1 Matching

For automatic extraction of image coordinates of tie points software *hwmatch1* is used. Originally, *hwmatch1* was developed at the FPF in Munich for frame images. The IPI implement in *hwmatch1* the extended functional model for three line imagery (Ebner et al., 1994) and modified the software according to the requirements of the Mars Express Mission (Heipke et al., 2004, Schmidt et al., 2005).

As input data the matching needs images, the observed exterior orientation, and the calibrated interior orientation parameters. As an optional input it is possible to use a MOLA DTM as approximate information.

The matching uses feature based techniques. Point features are extracted of the entire images using the Förstner operator. The images of all sensors are matched pairwise in all combinations using the cross correlation coefficient as similarity measure. The results of pixel correlations are sets of image coordinates of tie points for each image. In addition, the results are refined step by step through different levels of image pyramids. To optimize the results of lowest image pyramid, a Multi Least Image Squares Matching (MI-LSM) is used.

3.2 Mathematical model of bundle adjustment

In the bundle adjustment the concept of orientation points proposed by (Hofmann et al., 1982) is used. This approach estimates the parameters of the exterior orientation only at a few selected image lines, at the so-called orientation points. Additional DTM data as control information is used to fit photogrammetrically derived object points into the existing reference system on Mars.

3.2.1 Collinearity equations. The mathematical model for photogrammetric point determination with a three-line camera is based on the well known collinearity equations. These equations describe the fundamental geometrical condition that the rays through the three corresponding image points and the corresponding perspective centers intersect in the object point P_i (see Figure 3).

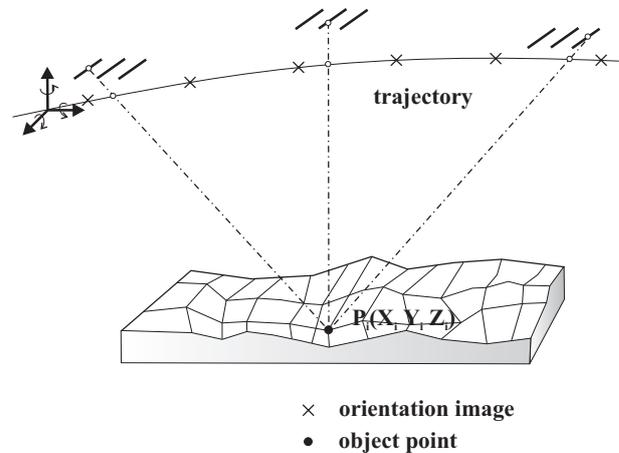


Figure 3: Imaging principle with three line camera

Two collinearity equations (Equation (1)) are established for each observed image coordinate. For every object point P_i there are several equations, because observed image coordinates are found in images of different sensors s .

$$\begin{aligned}
 x_{ijs} + \hat{x}_{0s} + \hat{v}_{x_{ijs}} &= \\
 &= c \frac{\hat{r}_{j11}(\hat{X}_i - \hat{X}_j) + \hat{r}_{j21}(\hat{Y}_i - \hat{Y}_j) + \hat{r}_{j31}(\hat{Z}_i - \hat{Z}_j)}{\hat{r}_{j13}(\hat{X}_i - \hat{X}_j) + \hat{r}_{j23}(\hat{Y}_i - \hat{Y}_j) + \hat{r}_{j33}(\hat{Z}_i - \hat{Z}_j)} \\
 & \quad (1) \\
 y_{ijs} + \hat{y}_{0s} + \hat{v}_{y_{ijs}} &= \\
 &= c \frac{\hat{r}_{j11}(\hat{X}_i - \hat{X}_j) + \hat{r}_{j21}(\hat{Y}_i - \hat{Y}_j) + \hat{r}_{j31}(\hat{Z}_i - \hat{Z}_j)}{\hat{r}_{j13}(\hat{X}_i - \hat{X}_j) + \hat{r}_{j23}(\hat{Y}_i - \hat{Y}_j) + \hat{r}_{j33}(\hat{Z}_i - \hat{Z}_j)}
 \end{aligned}$$

with

x_{ijs}, y_{ijs}	:	observed image coordinates of object point
$\hat{v}_{x_{ijs}}, \hat{v}_{y_{ijs}}$:	residuals of image coordinates
$\hat{X}_i, \hat{Y}_i, \hat{Z}_i$:	coordinates of object point P_i
$\hat{X}_j, \hat{Y}_j, \hat{Z}_j$:	coordinates of projective center
$\hat{r}_{j11}, \dots, \hat{r}_{j33}$:	elements of rotation matrix (Spiegel, 2007)
$\hat{x}_{0_s}, \hat{y}_{0_s}$:	unknown parameters of interior orientation
c	:	constant parameter of interior orientation

Generally, the collinearity equations are formulated for each pair of image coordinates. For this, it is necessary to improve the exterior parameters for each image line in which observed image coordinates are available. But, this is not possible because of geometric reasons. In the case of Mars Express (satellite Orbit) the trajectory is assumed to stable and the parameters of the exterior orientation are improved only at few selected positions m (orientation points). But, for all collinearity equations the exterior orientation parameters are needed. Therefore, the parameters of exterior orientation ($\hat{X}_j, \hat{Y}_j, \hat{Z}_j, \hat{\varphi}_j, \hat{\omega}_j, \hat{\kappa}_j$) laying between the orientation are formulated with Lagrange polynomials of grade three (Spiegel, 2007).

3.2.2 Observed unknowns. In the case of Mars Express, there are observed parameters of exterior orientation of each image line. At the orientation points m these observations can introduced in the adjustment with additional observation equations (Equation (2)).

$$\begin{aligned} \hat{v}_{X_m} &= \hat{X}_m - X_m \\ \hat{v}_{Y_m} &= \hat{Y}_m - Y_m \\ \hat{v}_{Z_m} &= \hat{X}_m - Z_m \\ \hat{v}_{\varphi_m} &= \hat{\varphi}_m - \varphi_m \\ \hat{v}_{\omega_m} &= \hat{\omega}_m - \omega_m \\ \hat{v}_{\kappa_m} &= \hat{\kappa}_m - \kappa_m \end{aligned} \quad (2)$$

with

$\hat{v}_{X_m}, \dots, \hat{v}_{\kappa_m}$:	residuals at exterior orientation points m
$\hat{X}_m, \dots, \hat{\kappa}_m$:	unknown exterior orientation parameters
	:	at orientation points m
X_m, \dots, κ_m	:	observed exterior orientation parameters
	:	at orientation points m

To avoid interpolation problems, constant differences ($\delta X_j, \delta Y_j, \delta Z_j, \delta \varphi_j, \delta \omega_j, \delta \kappa_j$) should be added to the exterior orientation parameters ($\hat{X}_j, \hat{Y}_j, \hat{Z}_j, \hat{\varphi}_j, \hat{\omega}_j, \hat{\kappa}_j$) at the image lines j between the orientation points (Spiegel, 2007).

3.2.3 Systematic effects in parameter of exterior orientation. The three-dimensional position of the spacecraft is determined by ESA up to 15 times per orbit. Because of doppler shift measurements there are systematic effects in observed exterior orientation (see Section 2.1). To modeling these effects in bundle adjustment it is possible to use additional bias (offset) and drift parameters. For this, it is necessary to expand equation (2) with bias and drift parameters (Equation (3)).

$$\begin{aligned} \hat{v}_{X_m} &= \hat{X}_m - X_m + \hat{X}_B + j \hat{X}_D \\ \hat{v}_{Y_m} &= \hat{Y}_m - Y_m + \hat{Y}_B + j \hat{Y}_D \\ \hat{v}_{Z_m} &= \hat{Z}_m - Z_m + \hat{Z}_B + j \hat{Z}_D \end{aligned} \quad (3)$$

with

$\hat{v}_{X_m}, \dots, \hat{v}_{Z_m}$:	residuals at exterior orientation points m
$\hat{X}_m, \dots, \hat{Z}_m$:	unknown exterior orientation parameters
	:	at orientation points m
X_m, \dots, Z_m	:	observed exterior orientation parameters
	:	at orientation points m
$\hat{X}_B, \dots, \hat{Z}_B$:	unknown bias
$\hat{X}_D, \dots, \hat{Z}_D$:	unknown drift
j	:	number of image line

Also, additional observation equations for bias and drift has to be introduced (Equation (4)).

$$\begin{aligned} \hat{v}_{X_B} &= \hat{X}_B - X_B \\ \hat{v}_{X_D} &= \hat{X}_D - X_D \\ \hat{v}_{Y_B} &= \hat{Y}_B - Y_B \\ \hat{v}_{Y_D} &= \hat{Y}_D - Y_D \\ \hat{v}_{Z_B} &= \hat{Z}_B - Z_B \\ \hat{v}_{Z_D} &= \hat{Z}_D - Z_D \end{aligned} \quad (4)$$

with

$\hat{v}_{X_B}, \dots, \hat{v}_{Z_B}$:	residuals of bias
$\hat{X}_B, \dots, \hat{Z}_B$:	unknown bias
X_B, \dots, Z_B	:	observed bias
$\hat{v}_{X_D}, \dots, \hat{v}_{Z_D}$:	residuals of drift
$\hat{X}_D, \dots, \hat{Z}_D$:	unknown drift
X_D, \dots, Z_D	:	observed drift

The attitude of the spacecraft is given by ESOC. But, only the predicted attitude are available (see Section 2.1). Because of this, it is not necessary to consider systematic effects in the attitude angles and to formulate additional equations.

3.2.4 Interior orientation. To adjust parameters of interior orientation, two observation equations per sensor are introduced in the bundle adjustment (Equation (5)).

$$\begin{aligned} \hat{v}_{x_{0_s}} &= \hat{x}_{0_s} - x_{0_s} \\ \hat{v}_{y_{0_s}} &= \hat{y}_{0_s} - y_{0_s} \end{aligned} \quad (5)$$

with

$\hat{v}_{x_{0_s}}, \hat{v}_{y_{0_s}}$:	residuals of parameter of interior orientation
$\hat{x}_{0_s}, \hat{y}_{0_s}$:	unknown parameters of interior orientation
x_{0_s}, y_{0_s}	:	observations of parameter of interior orientation
s	:	sensor

3.2.5 MOLA DTM as control information. Starting point of this discussion about DTM data as control information is the approach of (Strunz, 1993). This approach describes the use of DTM as additional or exclusive control information for aerial triangulation. Transferring this approach to the case of Mars Express and HRSC means that, the control information is the surface defined by MOLA DTM and HRSC points lie on these surfaces. A drawback of this approach is that it does not use the original MOLA track points but interpolated DTM points. The advantage

of this approach is that the effort to search for adequate neighboring MOLA points is reduced because the DTM has a regular grid structure.

This approach uses a least squares adjustment with additional conditions to get a relation between a DTM and the HRSC points. As already mentioned, the HRSC points have to lie on a bilinear surface defined by four neighboring MOLA DTM points, which enclose the HRSC point (see Figure 4). This condition can be formulated as a constraint on the vertical distance from the HRSC point to the bilinear surface. Furthermore, this constraint can be substituted by a fictive observation, used as additional observation in the bundle adjustment.

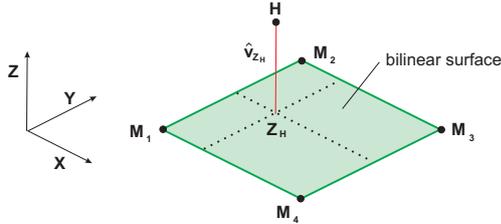


Figure 4: Fitting HRSC point in bilinear surface defined by MOLA DTM

The mathematical model has to be extended with an additional observation equation per object point (Equation (6)).

$$\hat{v}_{Z_H} = \hat{Z}_H - Z_H \quad (6)$$

with

- \hat{v}_{Z_H} : residuals of height \hat{Z}_H
- \hat{Z}_H : unknown height of object point H
- Z_H : fictive observation

The fictive observation Z_H for one object point is generated by bilinear interpolation of the MOLA surface (M_1, M_2, M_3, M_4) at the point of the unknown object points \hat{X}_H, \hat{Y}_H (Equation (7)).

$$\begin{aligned} Z_H = & \left[\left(1 - \frac{\hat{X}_H - X_{M_1}}{d}\right) \left(1 - \frac{\hat{Y}_H - Y_{M_1}}{d}\right) Z_{M_1} + \right. \\ & + \left(\frac{\hat{X}_H - X_{M_1}}{d} \right) \left(1 - \frac{\hat{Y}_H - Y_{M_1}}{d}\right) Z_{M_4} + \\ & + \left(1 - \frac{\hat{X}_H - X_{M_1}}{d}\right) \left(\frac{\hat{Y}_H - Y_{M_1}}{d} \right) Z_{M_2} + \\ & \left. + \left(\frac{\hat{X}_H - X_{M_1}}{d} \right) \left(\frac{\hat{Y}_H - Y_{M_1}}{d} \right) Z_{M_3} \right] \end{aligned} \quad (7)$$

with

- Z_H : fictive observation
- \hat{X}_H, \hat{Y}_H : unknown coordinates of object points H
- $X_{M_i}, Y_{M_i}, Z_{M_i}$: constant coordinates of MOLA DTM mesh points
- d : mesh width of MOLA DTM
($d = X_{M_4} - X_{M_1} = Y_{M_2} - Y_{M_1}$)

With this approach an improvement of the height (Z) can be expected, of course. An improvement in planimetry (X, Y) can only be determined, if there are different local terrain slopes at the different MOLA surfaces (Ebner and Ohlhof, 1994). Figure 5(a) shows a situation before bundle adjustment. After bundle adjustment, the differences between HRSC points and MOLA DTM are reduced (see Figure 5(b)).

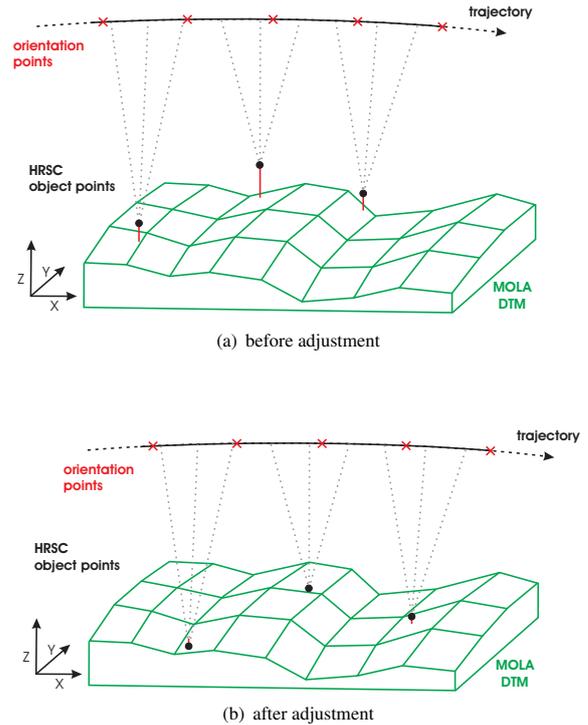


Figure 5: MOLA as control information

4 INTERIOR ORIENTATION

An investigation shows, that the parameters \hat{x}_{0_s} and \hat{y}_{0_s} of interior orientation can be improved for 7 of 9 lines. Therefore, two observation equations per sensor (=14 total) are introduced in the bundle adjustment. The other parameters of interior orientation are constant in adjustment.

For improvement of interior orientation, 46 orbits with good imaging conditions are selected. For all orbits the parameters \hat{x}_{0_s} and \hat{y}_{0_s} of nadir and the two stereo channels S1 and S2 are improved. For a part of selected orbits the color lines are not available in adequate geometric resolution. Therefore, the color lines can be improved only for 17 orbits. Afterwards the mean values for all orbits are computed and only one new set of interior orientation parameters is created.

Systematic residuals at the image coordinates of one orbit before adjustment of interior orientation is shown in figure 6. In the figure, each sensor is divided in 8 parts. For these parts the root mean square error is showed.

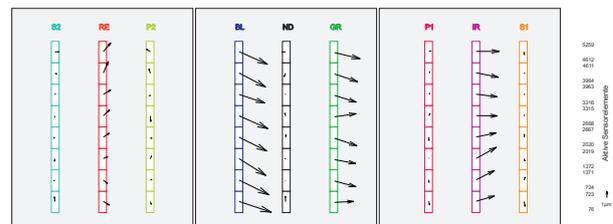


Figure 6: Residuals of image coordinates of orbit h2063 before adjustment of interior orientation

With new interior orientation, the residuals at image coordinates are reduced and there are no longer systematic residuals in the image coordinates (see Figure 7) compared to the case without new interior orientation.

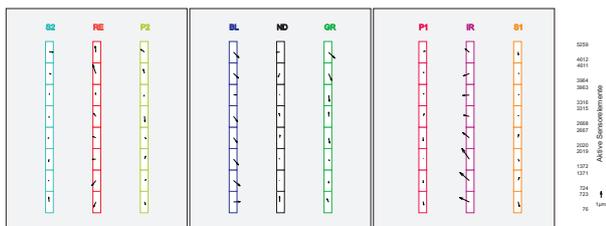


Figure 7: Residuals of image coordinates of orbit h2063 after adjustment of interior orientation

A comparison of Z-differences between HRSC object points and MOLA DTM shows the advantage of this approach. Certainly, the differences using calibrated interior orientation is better than without any bundle adjustment. But, the results can be improved with the new interior orientation.

5 RESULTS OF PROCESSING HRSC IMAGERY

In this section, the results of the bundle adjustment using the new interior orientation (considered to be stable) will be discussed for three cases: Using the nominal exterior orientation (Case A): Figure 8(a) shows the average displacements in planimetry between neighboring strips. The average displacements are in the range of 150 m up to 200 m in planimetry. The height differences between the HRSC object points and the MOLA DTM are irregularly distributed over the whole area (see Figure 8(b)).

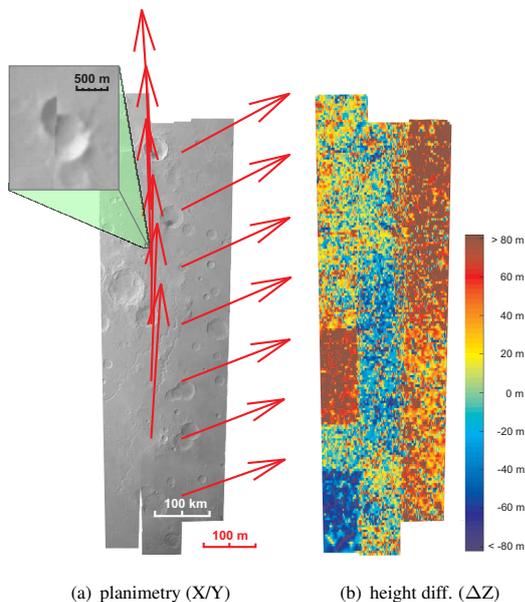


Figure 8: results before bundle adjustment

Improved exterior orientation with single strips (Case B): The Figure 9(a) shows the average displacements between strips, again. But, in this case the average displacements are smaller (only 15 m up to 40 m) as in case A. Also, the height differences are much smaller as in the case discussed before (see Figure 9(b)).

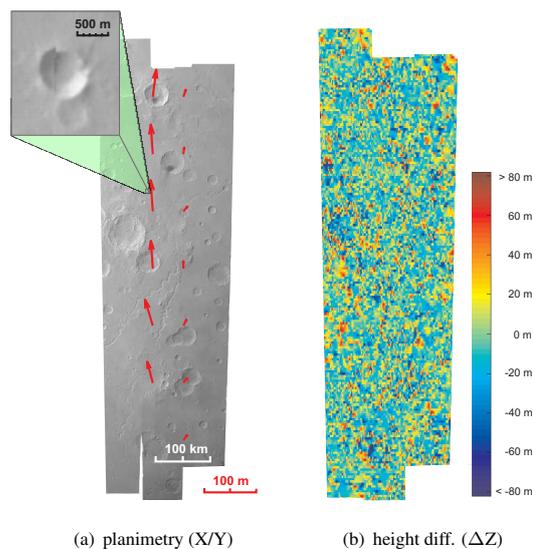


Figure 9: results of adjustment as single strip

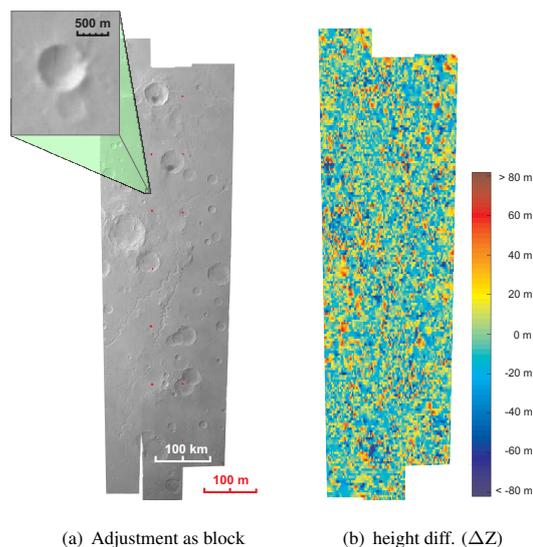


Figure 10: results of adjustment in a block

6 CONCLUSION

The parameter of exterior and interior orientation can be improved by using bundle adjustment. Regarding to interior orientation, a reduction of residuals at the image coordinates after bundle adjustment can be reached. Also, the Z-differences between HRSC object points and MOLA DTM is reduced for all 1200 investigated orbits with the new interior orientation, more or less. Therefore, it is absolutely necessary to use the new interior orientation in further investigations.

Improved exterior orientation adjusted in a block (Case C): The best results are achieved in this case. The average displacements in planimetry are lower than 10 m (Figure 10(a)). The improvement of the height differences between the HRSC object points and the MOLA DTM (see Figure 10(b)) is in this case only small because of the very good results in case B.

Using new interior orientation by computing single strips or blocks with bundle adjustment, the best result is reached with the bundle adjustment of a block. But also, the adjustment of single strips leads to good results. Finally, there are a high consistency between HRSC points and MOLA DTM, which constitutes the valid reference system on Mars.

REFERENCES

- Astrium, 2001. Mars Express – System Requirements Specification. Technical Report MEX.MMT.TN.0519, Astrium.
- Ebner, H. and Ohlhof, T., 1994. Utilization of Ground Control Points for Image Orientation without Point Identification in Image Space. In: *International Archives of Photogrammetry and Remote Sensing*, Vol. 30, Part 3/1, pp. 206–211.
- Ebner, H., Kornus, W. and Ohlhof, T., 1994. A Simulation Study on Point Determination for the MOMS-02/D2 Space Project using an Extended Functional Model. In: *Geo-Informationssysteme*, pp. 11–16.
- Ebner, H., Spiegel, M., Giese, A. B. B., Neukum, G. and the HRSC Co-Investigator Team, 2004. Improving the Exterior Orientation of Mars Express HRSC Imagery. In: *International Archives of Photogrammetry and Remote Sensing*, Vol. 35 (B4), pp. 852–857.
- Hechler, M. and Yáñez, A., 2000. Mars Express – Consolidated Report on Mission Analysis Issue 2.0. Technical Report MEX-ESC-RP-5500, ESA.
- Heipke, C., Schmidt, R., Oberst, J., Neukum, G. and the HRSC Co-Investigator Team, 2004. Performance of Automatic Tie Point Extraction Using HRSC Imagery of Mars Express Mission. In: *International Archives of Photogrammetry and Remote Sensing*, Vol. 35 (B4), pp. 846–851.
- Hofmann, O., Navé, P. and Ebner, H., 1982. DPS - A Digital Photogrammetric System for Producing Digital Elevation Models and Orthophotos by Means of Linear Array Scanner Imagery. In: *International Archives of Photogrammetry and Remote Sensing*, Vol. 24-III, pp. 216–227.
- Neukum, G. and Hoffmann, H., 2000. Images of Mars. In: R. Pellinen and P. Raudsepp (eds), *Towards Mars!*, pp. 129–152.
- Neumann, G., Lemoine, F., Smith, D. and Zuber, M., 2003. The Mars Orbiter Laser Altimeter Archive: Final Precision Experiment Data Record Release and Status of Radiometry. In: *Lunar and Planetary Science XXXIV*, Abstract #1978, Lunar and Planetary Institute, Houston (CD-ROM).
- Schmidt, R., Heipke, C., Brand, R., Neukum, G. and the HRSC Co-Investigator Team, 2005. Automatische Bestimmung von Verknüpfungspunkten in HRSC-Bildern der Mars Express Mission. In: *Photogrammetrie Fernerkundung Geoinformation*, E. Schweizerbart'sche Verlagsbuchhandlung, Stuttgart, pp. 373–379.
- Smith, D., Zuber, M., Frey, H., Garvin, J., Head, J., Muhleman, D., Pettengill, G., Phillips, R., Solomon, S., Zwally, H., Banerdt, W., Duxbury, T., Golombek, M., Lemoine, F., Neumann, G., Rowlands, D., Aharonson, O., Ford, P., Ivanov, A., Johnson, C., McGovern, P., Abshire, J., Afzal, R. and Sun, X., 2001. Mars Orbiter Laser Altimeter: Experiment Summary After the First Year of Global Mapping of Mars. In: *Journal of Geophysical Research*, Vol. 106(E10), pp. 23,689–23,722.
- Spiegel, M., 2007. Kombinierte Ausgleichung der Mars Express HRSC Zeilenbilddaten und des Mars Global Surveyor MOLA DGM. Technische Universität München, <http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:bvb:91-diss-20070801-618406-1-4>.
- Spiegel, M., Stilla, U., Giese, B., Neukum, G. and the HRSC Co-Investigator Team, 2005. Bündelausgleichung von HRSC-Bilddaten mit Mars Observer Laser Altimeter-Daten als Passinformation. In: *Photogrammetrie Fernerkundung Geoinformation*, E. Schweizerbart'sche Verlagsbuchhandlung, Stuttgart, pp. 381–386.
- Strunz, G., 1993. Bildorientierung und Objektrekonstruktion mit Punkten, Linien und Flächen. Dissertationsschrift, DGK-C, Nr. 408, Deutsche Geodätische Kommission, München.

ACKNOWLEDGEMENTS

The author thank the HRSC Experiment Teams at DLR Berlin and Freie Universität Berlin as well as the Mars Express Project Teams at ESTEC and ESOC for their successful planning and acquisition of data as well as for making the processed data available to the HRSC Team. The authors acknowledge the effort of Ralph Schmidt, Christian Heipke, Heinrich Ebner, Uwe Stilla, and Gerhard Neukum who have contributed to this investigation in the preparatory phase and in scientific discussions within the Team.

This work is funded by Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR) under grant no. 50 QM 0103. This support is gratefully acknowledged.

MULTI-VIEW STEREO AND LIDAR FOR OUTDOOR SCENE MODELLING

C. Strecha^a, W. von Hansen^b, L. Van Gool^{a,c} and U. Thoennessen^b

^a KU-Leuven ESAT/PSI, Belgium, {christoph.strecha,luc.vangool}@esat.kuleuven.be

^b FGAN-FOM Ettlingen, Germany {wvhansen,thoe}@fom.fgan.de

^c ETH Zürich BIWI, Switzerland

KEY WORDS: LIDAR , self-calibration, evaluation, multi-view stereo, Bayesian generative model based stereo

ABSTRACT:

In this paper we want to start the discussion on whether image based 3-D modelling techniques and especially multi-view stereo can possibly be used to replace LIDAR systems for outdoor 3D data acquisition. Two main issues have to be addressed in this context: (i) camera self-calibration and (ii) dense multi-view depth estimation. To investigate both, we have acquired test data from outdoor scenes with LIDAR and cameras. Using the LIDAR data as reference we provide an evaluation procedure to these two major parts of the 3D model building pipeline. The test images are available for the community as benchmark data.

1 INTRODUCTION

Several techniques to measure the shape of objects in 3-D are available. The most common systems are based on active stereo, passive stereo, time of flight laser measurements (LIDAR) or NMR imaging. For measurements in laboratories, one usually puts small objects *into* a relatively large measurement device. In contrast to this, outdoor measurements require to bring the measurement equipment *to* considerably larger objects. Traditionally, close range photogrammetry had been used to capture objects of architectural or archaeological interest such as buildings. This technique required precisely calibrated cameras as well as a lot of manual work on expensive coordinate measurement devices to produce a result. A second technique for precise outdoor measurements that has become important during the past years is the application of LIDAR systems. In contrast to image based techniques, these are able to directly produce a 3-D point cloud based on distance measurements with an accuracy of less than 1 cm. The downside are high costs for the system and – compared to taking images with a camera – data acquisition is time consuming.

Automatic reconstruction from multiple view imagery already is a low-cost alternative to traditional methods, but could even become a replacement to LIDAR systems once the geometrical accuracy of the results could be proven. For this purpose we have acquired LIDAR data and high resolution images for several outdoor scenes. The LIDAR data will serve as geometrical ground truth to evaluate the quality of the image based reconstructions.

To tackle the problem of 3-D data acquisition from images, one has to consider camera self-calibration techniques and multi-view stereo. Camera calibration includes the computation of external (position and orientation) and internal camera parameters. Multi-view stereo takes this information to compute a dense 3-D model of the scene. Both problems are usually seen separate such that we opted to provide evaluation data and an evaluation procedure to measure their accuracy independently. We consider this as an first step to the necessary evaluation which considers both problems in common¹. As an example we provide evaluation results for camera self-calibration (EPOCH, 2007) and a multi-view stereo approach that is based on generative models (Strecha et al., 2004, Strecha et al., 2006). The latter consists of two stages.

¹Working systems that automatically produce 3-D models from uncalibrated images are rarely available yet.

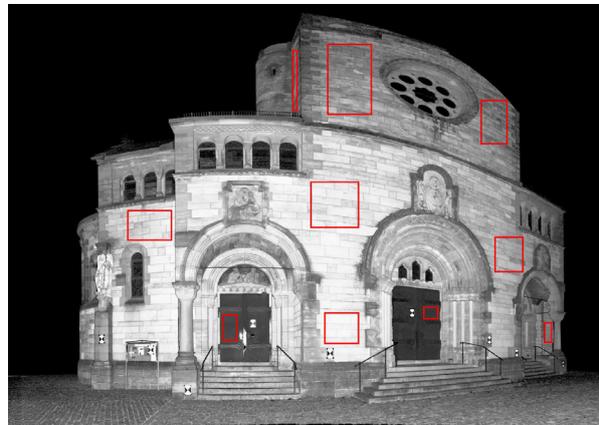


Figure 1: Example of the LIDAR data: intensity image of the church with marked planar regions.

Firstly, a global formulation considers all possible depth and visibility configurations of the scene (Strecha et al., 2006). This approach is relatively memory and time consuming and is therefore suited only for small resolution images. Its advantage is the ability to find a good, often global, solution without depth initialisation (Strecha, 2007). In a second stage, the solution of this global formulation is used as an initialisation to a PDE-based approach that is applied to large scale images (Strecha et al., 2004).

With this contribution we want to start the discussion on whether multi-view stereo can compete with LIDAR in terms of accuracy. Note, that we do not treat the LIDAR data as an *absolute* ground truth reference. The LIDAR itself contains errors, that are estimated and translated to the image based acquisition. As a result we evaluate image based depth modelling relative to the accuracy of the LIDAR data. This is different from previous ground truth data sets, where the reference data is assumed to be given without errors (Scharstein and Szeliski, 2002, Seitz et al., 2006, Bellmann et al., 2007). A second difference is the choice for outdoor scenes. Whereas indoor scenes can be measured by active stereo systems, *e.g.* by using structured light, this is not an option for outdoor environments. The evaluation of passive, image based 3-D modelling techniques is in that sense especially for outdoor scenes of large practical importance.

The paper is organised as follows. Sec. 2 deals with the LIDAR data acquisition, *i.e.* the registration of different scans and the ac-

Ambiguity interval	53.5 m	
Range resolution	0.8 mm	
Range noise	10 m 20%	3.0 mm rms
@ distance,	10 m 100%	1.3 mm rms
reflectivity	25 m 20%	9.0 mm rms
	25 m 100%	3.0 mm rms
Field of view ($v \times h$)	$310^\circ \times 360^\circ$	
Angular resolution	$0.018^\circ \times 0.01^\circ$	
Angular accuracy	$0.02^\circ \times 0.02^\circ$ rms	
Beam divergence	0.22 mrad	
Points per dataset	200 million	

Table 1: Specifications of the IMAGER 5003.

curacy estimation. We link the LIDAR point clouds to the images by estimating the camera parameters and depth maps together with all variances in sec. 3. In sec. 4 we consider the evaluation procedure and show example results. Sec. 5 concludes the paper.

2 ACQUISITION OF GROUND TRUTH DATA

2.1 Data acquisition and registration

The datasource for ground truth in our project is terrestrial laser scanning (LIDAR). A laser beam is scanned across the object surface, measuring the distance to the object for each position. Modern systems also record the returned intensity of the laser beam so that a gray scale texture image is available. We had a Zoller+Fröhlich IMAGER 5003 laser scanner at our disposition (tab. 1). Multiple scan positions are required for complex object surfaces to handle missing parts due to occlusions. Even though fully automatic methods exist for registration (Akca, 2003, Bae, 2006, Rabbani and van den Heuvel, 2005, von Hansen, 2006), we have chosen a semi-automatic way utilising software provided by the manufacturer in order to get a clean reference.

2-D targets are put into the scene and marked interactively in the datasets. Then the centre coordinates are automatically computed and a registration module computes a least squares estimation of the parameters for a rigid transform between the datasets. Poorly defined targets can be detected and manually removed. The resulting mean accuracy for a single target was 1.1 mm for the church and 1.5 mm for the fountain dataset. The targets are visible in the camera images as well and are used to link LIDAR and camera coordinate systems.

2.2 Preparation of point clouds

First, some filters from the software are applied to mask out bad points resulting from measurements into the sky, mixed pixels and other error sources. Some regions containing clutter have been manually masked out and a 3×3 median filter had been applied to the data to remove remaining noisy points.

Then, a triangle mesh has been created separately from each point cloud so that image rays can be intersected with object surfaces. The scan data is acquired in a system of lines and rows so that direct neighbors are known. Four points on a square can generate zero to two triangles. The decision is based on the angle α between the viewing ray and the normal vector of the triangle. In order to avoid the inclusion of depth edges as surfaces, triangles are only generated when $\cos \alpha > 0.5$. In the ambiguous case of two triangles, the configuration that minimises the larger of the two angles is chosen.

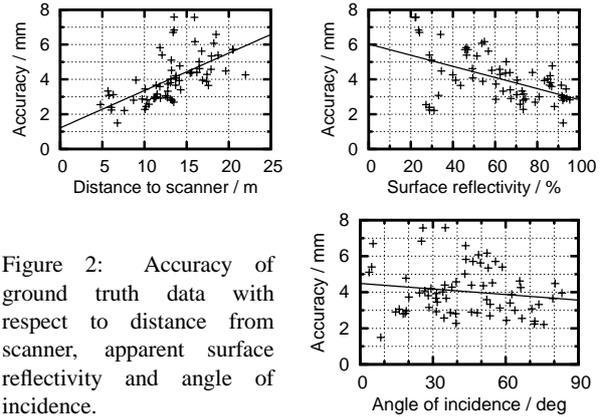


Figure 2: Accuracy of ground truth data with respect to distance from scanner, apparent surface reflectivity and angle of incidence.

2.3 Evaluation of point cloud accuracy

The manufacturer of the LIDAR system gives an accuracy of 1.3–9.0 mm for the range measurement based on distance and reflectivity (tab. 1). In this section, we will investigate the accuracy of our datasets. From scene knowledge, a total of 66 planar surface patches has been manually marked (fig. 1). For all the 3-D points inside each selection, a principal component analysis is performed. The covariance matrix of an ideal plane has rank two, for a real plane the smallest eigenvalue equals to the variance perpendicular to the plane. The associated eigenvector is the surface normal. As further parameters, the mean point of the surface as distance to the scanner and the mean gray value as estimate for the surface reflectivity are computed. Results are shown in fig. 2 separately for each parameter and are ordered by the amount of influence on the accuracy. The relationship between parameter and accuracy has been estimated by a linear function.

The most important factor is the distance from the scanner to the surface. For this reason, a maximum distance of 20–25 m has been used as cut-off during data export. The linear approximation ranges from 2–6 mm over the used range interval, yielding a variation of 4 mm. The apparent surface reflectivity is derived from the amount of light received by the sensor. These levels do not have a real physical meaning at this stage, but nevertheless covered the range from 20% to 95% without any adjustments. As with the distance, a relationship between reflectivity and accuracy exists, but the magnitude is slightly smaller. The linear approximation ranges from 3.0–5.5 mm, yielding only 2.5 mm variation. The last parameter is the angle of the surface normal with respect to the laser beam. An angle of 0° denotes a surface perpendicular to the light ray. This parameter is uncorrelated to the surface accuracy. One should be aware that the parameters are actually correlated, *e.g.* the apparent reflectivity depends on both distance and angle of incidence.

For all 66 regions chosen, the accuracy ranges from 1.5 mm for the best case to 7.6 mm for the worst case. The worst accuracies occur for regions on the church doors, which have a dark and reflective metal surface. Our results are in accordance to the accuracies given in tab. 1. From the diagrams an average accuracy of about 4 mm can be derived. Even though this value has been computed from planar regions, it can be generalised to any smooth object surface. In comparison to the accuracy of the targets (≤ 1.5 mm), we see that the registration error is smaller than the surface accuracy. It can therefore be concluded, that the registration is good enough as to not disturb the results.

The accuracy of 4 mm given here for the test data is an absolute accuracy. This is a high quality result as can be seen from the relative accuracy: Object size and typical measurement distances

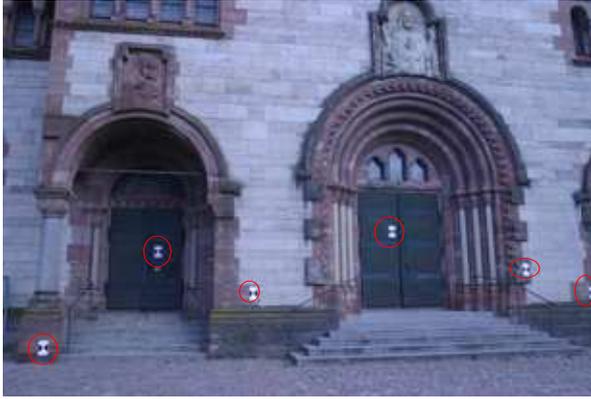


Figure 3: Example of target measurements with their (enlarged) covariance for the church data set.

are in the range of 15 m, leading to a relative accuracy of $3 \cdot 10^{-4}$. For comparison, the object size of Seitz *et al.* (Seitz et al., 2006) has been 0.1 m and the accuracy given was 0.25 mm, leading to a relative accuracy that is about ten times worse.

3 GROUND TRUTH ESTIMATION FROM LIDAR

Together with the LIDAR data the scenes have been captured with a Canon D60 digital camera with a resolution of 3072×2048 square pixels. In this section we describe how the images are calibrated and how the ground truth 3-D model is obtained by using the LIDAR data. Our focus is thereby not only on the ground truth estimation itself but also on the accuracy of our ground truth data. The LIDAR depth estimates are themselves the result of an erroneous measurement process and as such only given in a certain error range. We propagate these errors into our image based ground truth estimation.

Errors for the multi-view stereo evaluation are introduced by: (i) the 3-D accuracy of the LIDAR data itself and (ii) by the calibration errors of the input cameras. The latter does influence the quality of multi-view stereo reconstructions strongly. Evaluations that take these calibration errors into account should therefore be based on per image reference depth maps (more details are given in sec. 4.2) as opposed to Seitz *et al.* (Seitz et al., 2006), who evaluate the stereo reconstructions by the Euclidean 3-D distance between estimated and ground truth triangle mesh.

3.1 Mean and variance of the camera calibration

LIDAR data and camera images are linked via targets that are visible in both datasets. Thus the laser scanner provides 3-D reference coordinates that can be used to compute the calibration parameters for each camera. For the camera calibration we assume a perspective camera model with radial distortion (Hartley and Zisserman, 2000). The images are taken without changing the focal length, such that the internal camera parameters $\theta_{int} = \{f, s, x_0, a, y_0, k_1, k_2\}$ (\mathbf{K} -matrix and radial distortion parameters $k_{1,2}$) are assumed to be constant for all images. The external camera parameters are the position and orientation of the camera described by 6 parameters $\theta_{ext} = \{\alpha, \beta, \gamma, t_x, t_y, t_z\}$. The total number of parameters θ for N images is thus $7 + 6N$. To calibrate the cameras we used M targets, which have been placed in the scene. The 3-D position $\mathbf{Y}_j; j = 1 \dots M$ and the covariance Σ_Y for these is provided by the laser scan software. In each input image i we estimated the 2-D position \mathbf{y}_{ij} of these targets as it is shown in fig. (3) and used unit covariance Σ_{ij} .

Let \mathbf{y} denote all measurements, *i.e.* the collection of 3-D points \mathbf{Y}_j and the 2-D image measurements \mathbf{y}_{ij} . The expected value of the parameters $\theta = \{\theta_{int}, \theta_{ext1}, \dots, \theta_{extN}\}$ can be written as:

$$\mathbf{E}[\theta] = \int p(\mathbf{y}') p(\theta' | \mathbf{y}') \theta' d\mathbf{y}' d\theta'. \quad (1)$$

Here $p(\mathbf{y}')$ is the likelihood of data, *i.e.* among all 3-D points \mathbf{Y}'_i and image measurements \mathbf{y}'_{ij} only those will have a large likelihood that are close to the estimated values \mathbf{y} :

$$\begin{aligned} p(\mathbf{y}'_{ij}) &\propto \exp\left(-0.5(\mathbf{y}_{ij} - \mathbf{y}'_{ij})^T \Sigma_{ij}^{-1} (\mathbf{y}_{ij} - \mathbf{y}'_{ij})\right) \\ p(\mathbf{Y}'_j) &\propto \exp\left(-0.5(\mathbf{Y}_j - \mathbf{Y}'_j)^T \Sigma_Y^{-1} (\mathbf{Y}_j - \mathbf{Y}'_j)\right) \end{aligned} \quad (2)$$

The second term, $p(\theta | \mathbf{y}')$, is the likelihood of the calibration. This distribution is Gaussian and reflects the accuracy of the calibration, given the data points \mathbf{y}' . This accuracy is based on the reprojection error:

$$e(\theta) = \sum_i^N \sum_j^M (\mathbf{P}_i(\theta) \mathbf{Y}_j - \mathbf{y}_{ij})^T \Sigma_{ij}^{-1} (\mathbf{P}_i(\theta) \mathbf{Y}_j - \mathbf{y}_{ij}),$$

where $\mathbf{P}_i(\theta)$ projects a 3-D point \mathbf{Y}_j to the image pixel \mathbf{y}'_{ij} and the calibration likelihood becomes:

$$p(\theta | \mathbf{y}) \propto \exp(-0.5e(\theta)). \quad (3)$$

The covariance Σ of the camera parameters is similarly given by:

$$\Sigma = \int p(\mathbf{y}') p(\theta' | \mathbf{y}') (\mathbf{E}(\theta') - \theta') (\mathbf{E}(\theta') - \theta')^T d\mathbf{y}' d\theta'. \quad (4)$$

To compute the solution of eqn. (1) and (4), we apply a sampling strategy. The measurement distribution $p(\mathbf{y})$ is randomly sampled. Given a specific sample \mathbf{y}' the parameters θ' are computed by a bundle adjustment as the ML estimate of eq. (3):

$$\theta' = \arg \max_{\theta} \{\log p(\theta | \mathbf{y}')\}. \quad (5)$$

Using eq. (5) and eq. (2) we can approximate the expected values and the covariance in eqn. (1) and (4) by a weighted sum over the sample estimates. As a result we obtain all camera parameters θ by $\mathbf{E}[\theta]$ and their covariance Σ .

3.2 Mean and variance of ground truth depth

Given the mean and variance of the camera calibration we are now in the position to estimate the expected value of the per pixel depth and variance. Again we sample the camera parameter distribution given by $\mathbf{E}[\theta]$ and Σ in eq. (1) and eq. (4):

$$p(\theta') = \frac{\exp\left(-\frac{1}{2} (\mathbf{E}[\theta] - \theta')^T \Sigma^{-1} (\mathbf{E}[\theta] - \theta')\right)}{2\pi^{\frac{7+6N}{2}} |\Sigma|}, \quad (6)$$

and collect sufficient statistics for the per pixel depth values by the first intersection of the laser scan triangle mesh with the camera ray through each pixel. The result is the mean $\mathcal{D}_{\sigma}^{ij}$ and variance $\mathcal{D}_{\sigma}^{ij}$ of the depth value for all pixels i in all cameras j . An example is shown in fig. 4 for one camera. Note, that this procedure allows to evaluate multi-view stereo reconstructions independent on the accuracy of the camera calibration. If the performance of the stereo algorithm is evaluated in 3-D (*e.g.* by the Euclidean distance to the ground truth triangle mesh (Seitz et al., 2006)) the accuracy of the camera calibration and the accuracy of the stereo algorithm is mixed. Here, the evaluation is relative to calibration accuracy, *i.e.* pixels with a large depth variance, given

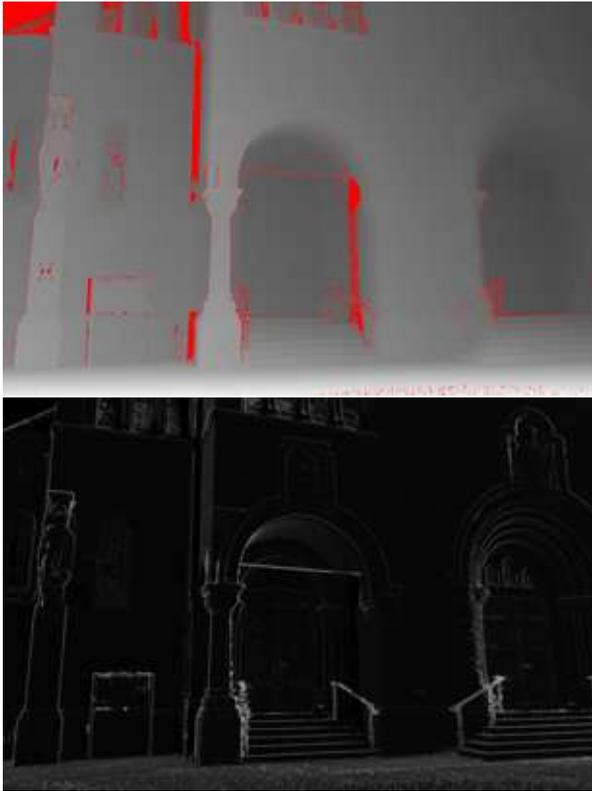


Figure 4: Mean $\mathcal{D}_{\text{LIDAR}}$ (top) and variance \mathcal{D}_σ (bottom) of the depth with respect to one camera. Red pixels denote pixels for which the camera ray does not intersect the laser scan model.

the uncertainty of the calibration, will influence the evaluation criterion accordingly. The variance plot in fig. 4 shows that large depth variance pixels appear near depth boundaries and for surface parts with a large slant. Obviously, these depth values vary most with a varying camera position. The reference depth maps (mean) and their variance will be used for evaluation in sec. 4.2.

4 DEPTH MODELLING FROM IMAGES

The 3-D modelling from high resolution images as the only input has made a huge step forward in being accurate and applicable to real scenes. Various authors propose a so called structure and motion pipeline (Akbarzadeh et al., 2006, Nister, 2004, Pollefeys et al., 2004, Rodriguez et al., 2005, Snavely et al., 2006, Strecha et al., 2003). This pipeline consists of mainly three steps. In the first step, the raw images undergo a sparse-feature based matching procedure. Matching is often based on invariant feature detectors (Mikolajczyk et al., 2005) and descriptors (Mikolajczyk and Schmid, 2005) which are applied to pairs of input images. Secondly, the position and orientation as well as the internal camera parameters are obtained by self-calibration (Hartley and Zisserman, 2000). The third step takes the input images, which have often been corrected for radial distortion, and the camera parameters and establishes dense correspondences or the complete 3-D model (see (Seitz et al., 2006) for an overview). In the remainder of this section we discuss and evaluate self-calibration and multi-view stereo on these ground truth data (fig. 3 and fig. 5).

4.1 Camera calibration

To compare results of self-calibration techniques based on our ground truth data we first have to align both camera tracks by a



Figure 5: Example image from the fountain data set.

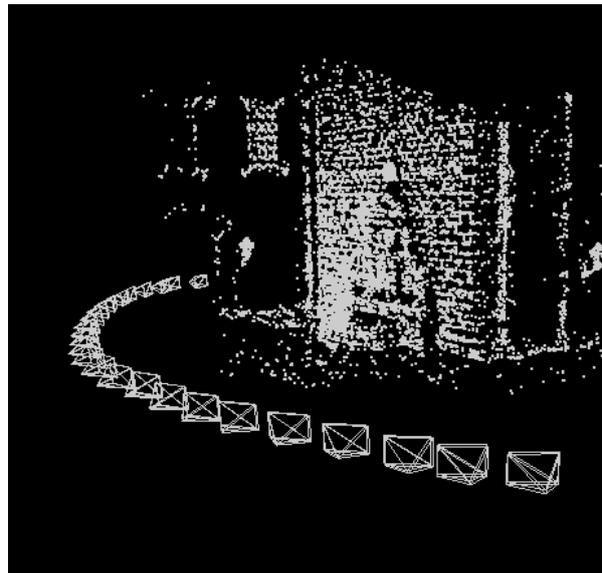


Figure 6: Camera calibration for the fountain data set in fig. 5

rigid 3-D transformation (scale, rotation and translation). First the scale is estimated by using the largest camera distance within each camera track. For one camera we estimate rotation and translation. These values are used as an initialisation to a non-linear optimisation that brings the evaluation track into the coordinate system of the ground truth track. After alignment the error of a certain calibration θ_{eval} is given by the negative log-likelihood of eq. (6):

$$\epsilon = (\mathbf{E}[\theta] - \theta_{eval})^T \Sigma^{-1} (\mathbf{E}[\theta] - \theta_{eval}). \quad (7)$$

As an example we tested the calibration with the EPOCH (EPOCH, 2007) software which successfully calibrated all cameras for the fountain data set. The result of this automatic camera calibration is shown in fig. 6, where we plot the position and orientation of the cameras (both ground truth and the EPOCH cameras). Fig. 7 shows the difference in (x,y,z) position for each of the 25 cameras w.r.t. the ground truth position. The error bars indicate the 3σ value of our ground truth camera positions. Note, the larger variance for the first camera positions in this data set. They are due to a larger distance to the scene which results in less reliable camera positions. The camera positions lie within one cm from the ground truth and the average distance is 5 times larger than the position variance of the ground truth.

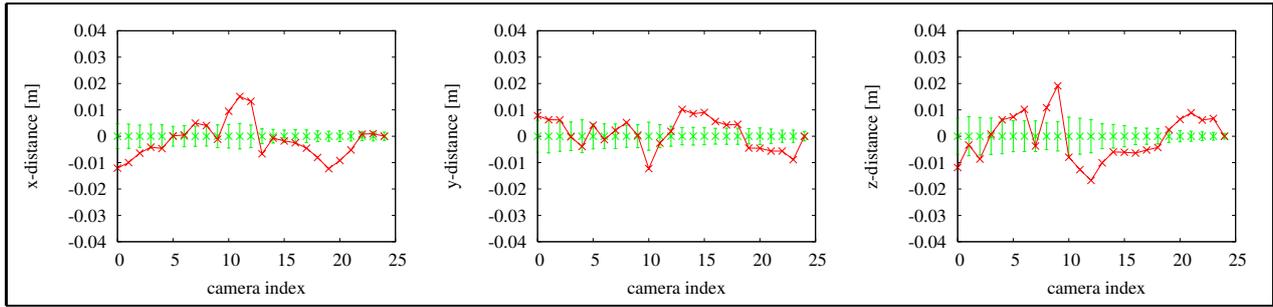


Figure 7: Position error [m] of the self-calibration (red) for the fountain data set in fig. 5. The green error bars indicate the 3σ value of the ground truth positions.

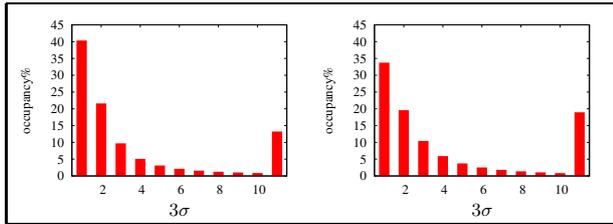


Figure 8: Histogram of the depth errors for the fountain (left) and church data set (right).

4.2 Multi-view stereo

Dense multi-view stereo applied to outdoor scenes cannot rely on visual hulls that are very useful for indoor stereo applications as evaluated by Seitz *et al.* (Seitz et al., 2006). Our test images have high resolution in order to meet the LIDAR precision and do not capture the object from all around. During data acquisition we also encountered problems due to pedestrians and changing light conditions. These aspects form a particular challenge for outdoor stereo reconstructions.

As input to the multi-view stereo algorithms we provide the ground truth camera calibration as estimated in sec. 3, the images, which have been corrected for radial distortion, the bounding volume of the scene as well as the minimal/maximal depth value w.r.t. to each input image. We evaluate the results for our generative model based multi-view stereo approach as formulated in (Strecha et al., 2004, Strecha et al., 2006, Strecha, 2007).

The results of this multi-view stereo approach are shown in fig. 9 for the fountain data. The accuracy of the stereo reconstruction $\mathcal{D}_{Stereo}^{ij}$ is evaluated by building a histogram h_k over the relative errors:

$$h_k \propto \sum_{ij} \delta_k \left(\left| \mathcal{D}_{LIDAR}^{ij} - \mathcal{D}_{Stereo}^{ij} \right|, \mathcal{D}_{\sigma}^{ij} \right). \quad (8)$$

\mathcal{D}_{LIDAR}^{ij} is the expected value of the LIDAR depth estimate at pixel position i and camera j and $\mathcal{D}_{\sigma}^{ij}$ its corresponding variance (both shown in fig. 4). Furthermore, $\delta_k(\cdot)$ is an indicator function which evaluates to 1 if the depth difference $\left| \mathcal{D}_{LIDAR}^{ij} - \mathcal{D}_{Stereo}^{ij} \right|$ falls within the variance range $[k\mathcal{D}_{\sigma}^{ij}, (k+1)\mathcal{D}_{\sigma}^{ij}]$ and evaluates to zero otherwise. The stereo estimate $\mathcal{D}_{Stereo}^{ij}$ is obtained from a 3-D triangle mesh by computing the depth of the first triangle intersection with the j^{th} camera ray going through pixel i . Its value is given directly for multi-view stereo formulations that use a depth-map representation (Seitz et al., 2006). All depth estimates for which the absolute difference with the ground truth is larger than $30\mathcal{D}_{\sigma}^{ij}$ and all pixels for which the multi-view stereo

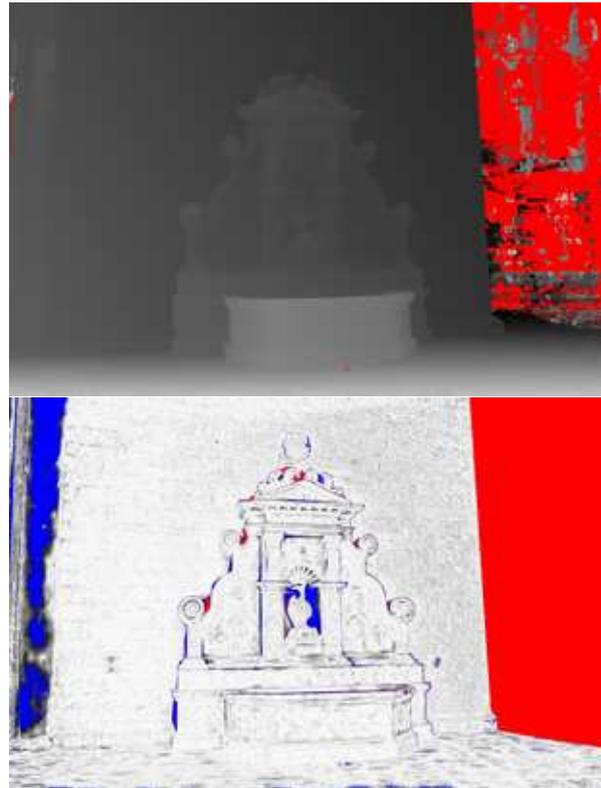


Figure 9: Stereo depth-map (top) and the variance weighted depth difference w.r.t. to the LIDAR ground truth (bottom): white pixels indicate small relative error; missing ground truth depth values are in red; blue pixels denote errors for which $\left| \mathcal{D}_{LIDAR} - \mathcal{D}_{Stereo} \right| / \mathcal{D}_{\sigma} \geq 30$.

reconstruction does not give depth estimates are collected all together in the last bin. These are all pixels indicated by blue in fig. 9. The relative error histogram for the fountain and church data are show in fig. 8. They can be interpreted as follows: $\sim 40\%$ ($\sim 34\%$) of the stereo depth estimates lie within the 3σ range of the LIDAR data for the fountain (church) data set; for $\sim 15\%$ ($\sim 20\%$) either no stereo depth exists or the error is larger than 30 times the LIDAR variance. Note, that our evaluation is relative to the camera calibration accuracy. This is achieved by evaluating the depth accuracy in the image by a local, pixel dependent value of $\mathcal{D}_{\sigma}^{ij}$. Untextured renderings of the multi-view stereo and LIDAR models are shown in fig. 10.

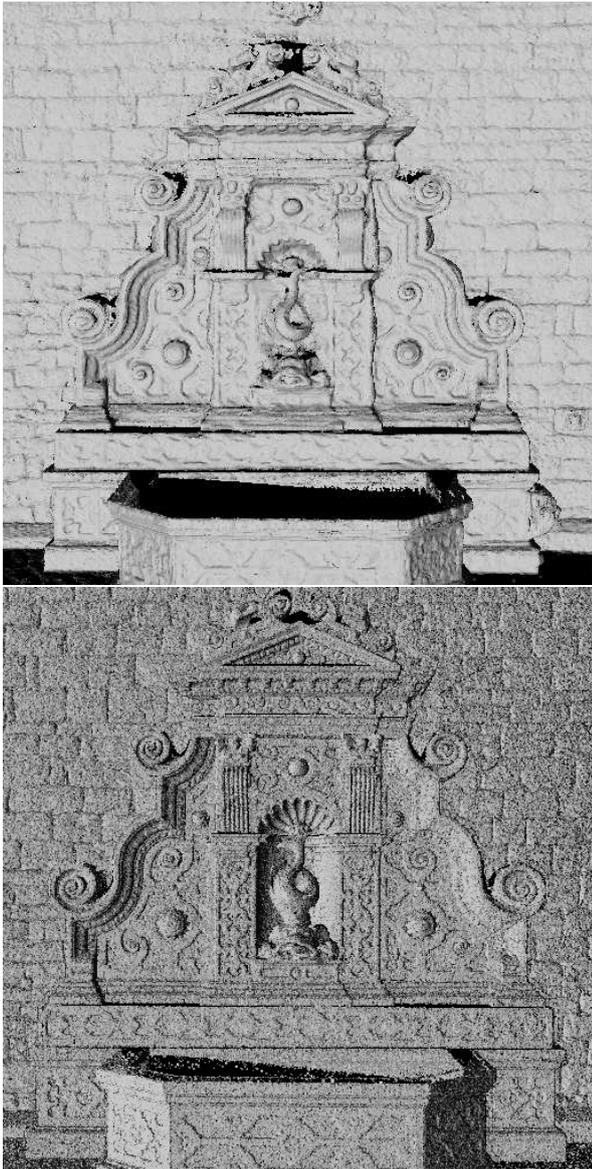


Figure 10: Multi-view stereo (top) and LIDAR (bottom) renderings of the fountain scene.

5 SUMMARY AND CONCLUSIONS

In this paper we investigated the possibility to evaluate 3-D modelling of outdoor scenes based on high resolution digital images using ground truth acquired by a LIDAR system. An evaluation carried out on the LIDAR data showed an average geometric accuracy of the surface measurements of about 4 mm and less than 1.5 mm. This is slightly better than the accuracy achieved from the images so that we can therefore conclude that our ground truth is of sufficient quality for further tests.

The link between LIDAR point cloud and images has been established via targets visible in both datasets and by generating per pixel depth maps with associated accuracies for each of the images. This allows to take into account the individual camera configuration so that poorly defined 3-D points will not deteriorate the evaluation result.

Example results have been shown for the the two important parts of image based 3-D acquisition, *i.e.* camera self-calibration and

multi-view stereo. The reference data has been computed such that each of these steps can be evaluated separately and relative to the accuracy of the reference data. We will make the images available as benchmark data in order to help to enhance the accuracy of automatic image based 3-D modelling techniques. We hope that vision systems become available that can compete with LIDAR systems in terms of geometric quality.

REFERENCES

- Akbarzadeh, A., Frahm, J., Mordohai, P., Clipp, B., Engels, C., Gallup, D., Merrell, P., Phelps, M., Sinha, S., Talton, B., Wang, L., Yang, Q., Stewenius, H., Yang, R., Welch, G., Towles, H., Nister, D. and Pollefeys, M., 2006. Towards urban 3D reconstruction from video. In: *Proc. 3DPVT*, pp. 1–8.
- Akca, D., 2003. Full automatic registration of laser scanner point clouds. In: *Optical 3-D Measurements VI*, pp. 330–337.
- Bae, K.-H., 2006. Automated Registration of Unorganised Point Clouds from Terrestrial Laser Scanners. PhD thesis, Curtin University of Technology.
- Bellmann, A., Hellwich, O., Rodehorst, V. and Yilmaz, U., 2007. A benchmarking dataset for performance evaluation of automatic surface reconstruction algorithms. In: *ISPRS Workshop BenCOS*.
- EPOCH, 2007. www.esat.kuleuven.be/~visit3d/webservice/html.
- Hartley, R. and Zisserman, A., 2000. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521623049.
- Mikolajczyk, K. and Schmid, C., 2005. A performance evaluation of local descriptors. *IEEE Transactions on PAMI* 27(10), pp. 1615–1630.
- 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. *IJCV* 65(1-2), pp. 43–72.
- Nister, D., 2004. Automatic passive recovery of 3D from images and video. In: *Proc. 3DPVT*, IEEE Computer Society, Washington, DC, USA, pp. 438–445.
- Pollefeys, M., Van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J. and Koch, R., 2004. Visual modeling with a hand-held camera. *IJCV* 59(3), pp. 207–232.
- Rabbani, T. and van den Heuvel, F., 2005. Automatic point cloud registration using constrained search for corresponding objects. In: *7th Conference on Optical 3-D Measurements*.
- Rodriguez, T., Sturm, P., Gargallo, P., Guilbert, N., Heyden, A., Menendez, J. and Ronda, J., 2005. Photorealistic 3D reconstruction from hand-held cameras. *Machine Vision and Applications* 16(4), pp. 246–257.
- Scharstein, D. and Szeliski, R., 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV* 47(1/2/3), pp. 7–42.
- Seitz, S., Curless, B., Diebel, J., Scharstein, D. and Szeliski, R., 2006. A comparison and evaluation of multi-view stereo reconstruction algorithms. In: *Proc. CVPR*, IEEE Computer Society, Washington, DC, USA, pp. 519–528.
- Snaveley, N., Seitz, S. and Szeliski, R., 2006. Photo tourism: exploring photo collections in 3D. In: *SIGGRAPH '06*, ACM Press, New York, NY, USA, pp. 835–846.
- Strecha, C., 2007. Multi-view stereo as an inverse inference problem. PhD thesis, PSI-Visics, KU-Leuven.
- Strecha, C., Fransens, R. and Van Gool, L., 2004. Wide-baseline stereo from multiple views: a probabilistic account. *Proc. CVPR* 1, pp. 552–559.
- Strecha, C., Fransens, R. and Van Gool, L., 2006. Combined depth and outlier estimation in multi-view stereo. *Proc. CVPR* pp. 2394–2401.
- Strecha, C., Tuytelaars, T. and Van Gool, L., 2003. Dense matching of multiple wide-baseline views. In: *Proc. ICCV*, pp. 1194–1201.
- von Hansen, W., 2006. Robust automatic marker-free registration of terrestrial scan data. In: W. Förstner and R. Steffen (eds), *Photogrammetric Computer Vision*, IAPRS, Vol. XXXVI Part 3.

USING PAVEMENT MARKINGS TO SUPPORT THE QA/QC OF LIDAR DATA

C. Toth^{a,*}, E. Paska^a, D. Brzezinska^b

^a Center for Mapping, OSU, 1216 Kinnear Road, Columbus, OH 43212 USA - (toth, eva-paska)@cfm.ohio-state.edu

^b Dept. of Civil and Environmental Engineering and Geodetic Science, OSU, dbrzezinska@osu.edu

Commission I, WG I/2

KEY WORDS: LiDAR, LiDAR intensity, Feature extraction, QA/QC

ABSTRACT:

LiDAR technology became an indispensable airborne mapping tool in recent years and is the primary source of highly accurate surface data at large scale. Although, the ranging accuracy of the laser sensor strongly depends on the surface characteristics, by and large, it falls in to the few cm range. This also implies that the achieved accuracy of a LiDAR system, defined in terms of the absolute accuracy of the laser points, is predominantly determined by the quality of the navigation solution (typically based on GPS/IMU sensor integration). Despite significant advancements in navigation technologies recently, to achieve and sustain a high accuracy navigation solution of an airborne platform for extended time is still a difficult task. Furthermore, there is no reliable way to assess the positioning quality of the data captured by any imaging sensor systems, which are based on direct georeferencing. Therefore, using some ground control is almost mandatory if high accuracy is required. This paper introduces a method to use road pavement marking as ground control that could be used for QA/QC. These linear features are widely available in urban areas and along transportation corridors, where most of the government and commercial mapping takes place. A key advantage of using pavement markings is that they can be quickly surveyed with GPS VRS technique.

1. INTRODUCTION

The introduction of airborne LiDAR (Light Detection And Ranging) in the late nineties was followed by a quick proliferation of the technology, and LiDAR is now the primary surface data extraction mapping technique. This remarkable success is mainly due to the fact that LiDAR data are explicit and the processing can be highly automated plus that an unprecedented vertical accuracy could be obtained relatively easily. The horizontal accuracy of the LiDAR data was not a concern in the early use of this technology. In fact, the first LiDAR data QA/QC and product characterization effort did only deal with the vertical accuracy (ASPRS, 2004).

As the LiDAR market started to grow rapidly, soon the LiDAR systems showed truly phenomenal performance improvements. In less than five years, the pulse rate improved by an order and now 100 and 150 kHz systems are widely used (Optech, 2006 and Leica, 2006) and experimental two-pulse systems are also available. More importantly, the ranging accuracy has increased substantially and now stands close to the level of static GPS surveys, i.e., 1-2 cm for hard surfaces, which is practically negligible to the typical navigation error budget. This remarkable performance potential of the newer LiDAR systems, combined with better operational techniques, opened the door toward applications where large-scale or engineering-scale accuracy is required. At this point the georeferencing error budget and, to a lesser extent, the sensor calibration quality, are critical to achieving engineering design level accuracy (few cm). Using ground control is an effective way to compensate for georeferencing and sensor modeling errors. In addition, ground control can provide for independent and highly reliable QA/QC processes.

This paper proposes a method to use road pavement markings as ground control to assess the quality of the LiDAR data as well as to improve the point cloud accuracy by post-processing. Beyond their wide availability, the use of pavement markings is primarily motivated by the fact that they can be rather easily surveyed using GPS VRS (Virtual Reference System) technology; the process is fast, typically it takes one minute to survey a point, and the accuracy, in general, is about 2-3 and 3-6 cm horizontally and vertically, respectively.

2. LIDAR ACCURACY AND ERROR CORRECTION TECHNIQUES

The errors in laser scanning data can come from individual sensor calibration or measurement errors, lack of synchronization, or misalignment between the different sensors. Baltsavias (1999) presents an overview of the basic relations and error formulae concerning airborne laser scanning. Schenk (2001) provides a summary of the major error sources for airborne laser scanners and error formulas focusing on the effect of systematic errors on point positioning. In general, LiDAR sensor calibration includes scan angle, range calibration and intensity-based range correction. The LiDAR sensor platform orientation is always provided by a GPS/IMU-based integrated navigation system. The connection between the navigation and LiDAR sensor frames is described by the mounting bias, which is composed of the offset between the origin of the two coordinate systems and the boresight misalignment (the boresight misalignment describes the rotation between the two coordinate systems, and is usually expressed by roll, pitch and heading angles). To achieve optimal error compensation that assures the highest accuracy of the final product, all of these parameters should be calibrated. Since not all of the parameters can be calibrated in a laboratory

* Corresponding author.

environment, a combination of laboratory and in situ calibrations is the only viable option for LiDAR system calibration. Typical anomalies in the LiDAR data indicating system calibration errors are: edges of the strips could bend up or down (scan angle error), horizontal surfaces have a visible mismatch between the known and the LiDAR point-defined surfaces (boresight misalignment or navigation error), vertical coordinates of LiDAR points over flat areas do not match the known vertical coordinate of the area (ranging or navigation error), objects, such as pavement markings made of retro reflective coatings, may show up above the surface level, although they should practically have identical vertical coordinates (lack of intensity correction of the range data), etc.

The techniques to detect and ultimately compensate for errors fall into two broad categories based on whether they use absolute control or not. The first group includes most of the strip adjustment techniques and some of the sensor and boresight calibration methods. The ground control-based techniques encompass comparisons to reference surfaces, such as parking lots and buildings, and methods using LiDAR-specific control targets.

Strip adjustment methods primarily minimize the vertical discrepancies between overlapping strips or between strips and horizontal control surfaces. These strip adjustments can be referred to as one-dimensional strip adjustment methods (Crombaghs et al., 2000; Kager and Kraus, 2001); tie or absolute control features used for this adjustment are flat horizontal surfaces. The problem with this kind of adjustment is that existing planimetric errors are likely to remain in the data. Vosselman and Maas (2001) have shown that systematic planimetric errors are often much more significant than vertical errors in LiDAR data and, therefore, a 3D strip adjustment is the desirable solution minimizing the 3D discrepancies between overlapping strips and at control points. A number of 3D adjustment methods have been published. Kilian et al. (1996) presented a method of transforming overlapping LiDAR strips to make them coincide with each other using control and tie points in a way similar to photogrammetric block adjustment. Burman (2002) treated the discrepancies between overlapping strips as positioning and orientation errors with special attention given to the alignment error between the IMU and laser scanner (Soininen, 2005). Filin (2003) presented a similar method for recovering the systematic errors; the method is based on constraining the position of the laser points to the surface from which it was reflected. Toth et al. (2002) presented a method that tried to make overlapping strips coincide, with the primary objective of recovering the boresight misalignment between the IMU and laser sensor.

LiDAR-specific ground control targets were introduced by Toth and Brzezinska (2005; Csanyi et al., 2005). The proposed technique uses ground control targets specifically designed for LiDAR data to provide quality control for applications that require cm-level, engineering scale mapping accuracy. Simulation results confirmed that the optimal target is rotation invariant, circular-shaped, elevated from the ground and that a flat target with 1 m circle radius can provide sufficient accuracy from a point density of about 5 pts/m². Targets larger than 2 m in diameter will not lead to significant improvements. In addition, a two-concentric-circle design (the inner circle has one-half the radius of the outer circle) with different coatings can produce considerable accuracy improvements in the horizontal position. Details and performance evaluation can be found in (Csanyi and Toth, 2007).

3. LIDAR INTENSITY DATA

The introduction of intensity data a few years ago produced unrealistically high initial expectations. On one side, the visualization value provided a major help for interactive processing, and thus, users could immediately benefit from the new source of data, as LiDAR intensity was comparable to optical image type of data that had been missed by practitioners from the early beginning. On the other side, the algorithmic advantages of using intensity data for providing better LiDAR data processing were largely overestimated. While research instantly started to address the exploitation of the new source of information, the problem seemed to be harder than expected. In simple terms, the major difficulty of working with LiDAR intensity data is the relative nature of this signal. For example, different surfaces, data from different flying heights, and different surface orientations can produce exactly the same intensity values. Therefore, techniques to calibrate the intensity and range values with respect to each other started to become more common.

One of the first attempts on using intensity data dates back to the time when LiDAR intensity data were not yet commercially available. Maas (2001) describes the extension of a TIN-based matching technique using reflectance data (LiDAR intensity data) to replace surface height texture for the determination of planimetric strip offsets in flat areas with sufficient reflectance texture. As an extension, Vosselman (2002) offers another solution, kind of a feature-based matching, to avoid interpolation of the data, using linear features, gable roofs, and ditches, modeled by analytical functions that can provide accurate offset determination. Later, research interest steered toward conventional classification use of the intensity data. Song et al. (2002) proposed a technique to use intensity data for land-cover classification. A similar study on using intensity for glacier classification is presented in Lutz et al. (2003). A recent review of more advanced versions of these techniques is offered by Hasegawa (2006). A comprehensive study on processing both range and intensity data is provided by Sithole (2005). Kaasalainen et al. (2005) provides a review on intensity data with respect to calibration. Nobrega and O'Hara (2006) compare two techniques for filtering intensity data for object extraction. Finally, Ahokas et al. (2006) presents the results of a calibration test on intensity data using the Optech ALTM 3100.

Figures 1 and 2 show simultaneously acquired orthoimage and the LiDAR intensity image, respectively, of an intersection. The LiDAR point density was about 4 pts/m² with foot print size of 15 cm. Note that the pavement markings in the LiDAR image are quite visible and distinct from the pavement. Consequently, if the approximate location of the pavement markings is known, then their extraction is a fairly straightforward task.

To illustrate that LiDAR elevation and intensity data are correlated and intensity information can indicate the presence of ranging error, Figure 3 shows the elevation data of the same intersection. Note that the pavement markings can be seen quite well, which conflicts with the fact that elevation value of the markings and the pavement around them should be identical (the few mm thickness of the markings is negligible compared to the few cm ranging accuracy of the laser system). This phenomenon is known and correction tables are available to partially compensate for this effect. The importance of this anomaly from our perspective is that during the comparative analysis later, the elevation value of the markings should be replaced by the average elevation of the pavement.

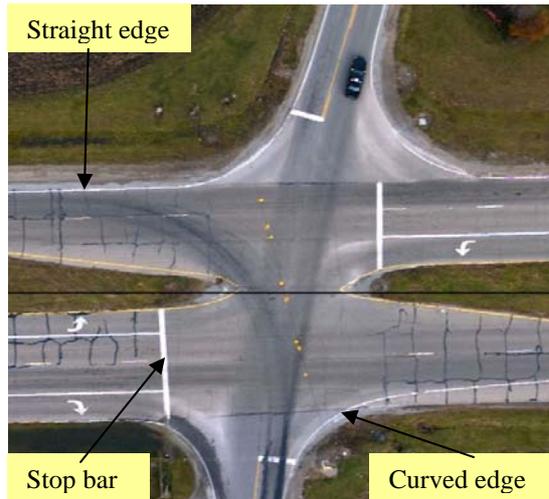


Figure 1. Typical pavement markings at an intersection.



Figure 2. LiDAR intensity image.

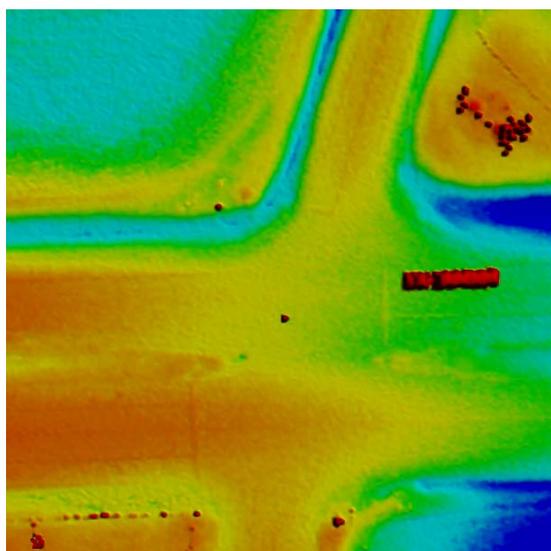


Figure 3. LiDAR elevation data.

4. EXTRACTING PAVEMENT MARKINGS AND USING THEM AS GROUND CONTROL

The concept of the proposed method, including pavement marking extraction as well as parameterization of the marks based on LiDAR intensity data, the comparison with ground truth, and the determination of a transformation to correct the point cloud, analysis of result, etc., is shown in Figure 4. General assumptions are that the survey data of the pavement markings are available a priori and the individual point accuracy, describing the marks, is known at the cm-level. To achieve good performance, sufficient number of pavement markings is required with good spatial distribution. At this point only three types of pavement markings are considered: Stop bars, straight edge lines and curved edge lines; Figure 1 shows the three pavement marking types. The survey data of the pavement markings is provided as point observations along the centerline of the markings. The LiDAR data, including range and intensity components, are assumed to be of reasonable quality; i.e., the point cloud accuracy is better than a meter.

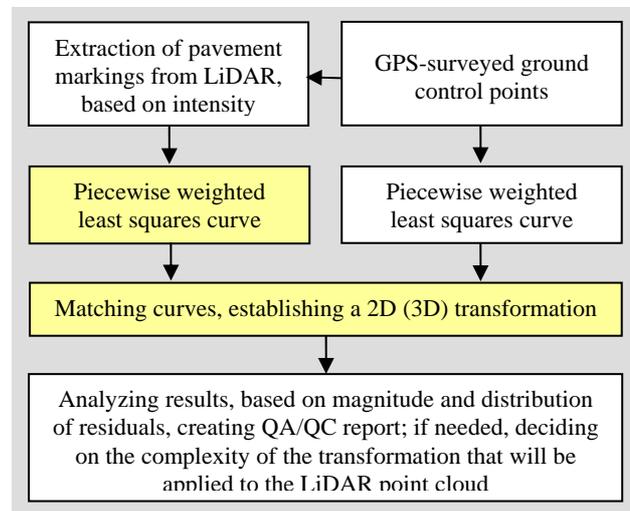


Figure 4. Block diagram of the proposed method.

Based on the comparison of the two descriptions of pavement markings, one obtained from the GPS survey and the other one from LiDAR intensity data, offset and orientation differences can be detected. Depending on the magnitude of the observed differences and their spatial distribution, a variety of corrections can be applied to the LiDAR point cloud to improve the point position accuracy. For example, if there is a similar vertical shift detected at the control features, a common vertical offset correction can be applied. If the amount of vertical shift detected varies by location and/or combined with non negligible horizontal differences, a more complex model, such as a 3D similarity transformation can be applied. Note that assessing the horizontal accuracy is difficult, as it is mainly defined by the footprint of the laser pulse, which depends on flying height and beam convergence; in addition, the impact of object surface characteristics could be also significant. The transformation based on the observed differences can be formulated on both, point- and linear feature-based least squares adjustment techniques. The conventional control point-based method is rather straightforward; similar to an absolute orientation of a stereo model with fixed scale. Linear feature-based orientation is less widely used, but could be feasible given the availability of matched linear features. Finally, if the differences are out of

the usual range (gross errors), the process can indicate system malfunctioning.

In our case, the point-based transformation is directly not applicable, as there is no point-to-point correspondence between the two point sets that describe the same linear feature. Assuming that the two representations provide an adequate description of the same shape, the problem is simply how to match two free-shape curves. In the following, the two key components of the proposed method, curve fitting and matching are discussed at detail.

4.1 Curve fitting

The extracted LiDAR points of the pavement markings and their surveyed data should be modeled as linear features in order to be matched with each other. The selected method is an extended version of the algorithm, originally proposed by Ichida and Kiyono in 1977, and is a piecewise weighted least squares curve fitting based on cubic (third-order polynomial) model, which seemed to be adequate for our conditions. To handle any kind of curves, defined as the locus of points $f(x, y) = 0$ where $f(x, y)$ is a polynomial, the curve fitting is performed for smaller segments in local coordinate systems, which are defined by the end points of the curve segments. The primary advantage of using a local coordinate system is to avoid problems when curves become vertical in the mapping coordinate system. Figure 5 shows the concept of the local coordinate system used for curve fitting; obviously, the fitting results as well as the fitting constraints are always converted forth and back between the local and mapping coordinate frames.

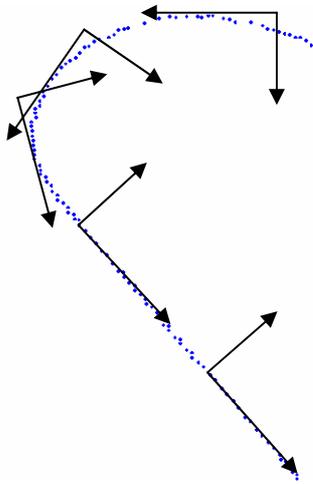


Figure 5. The curve fitting is done in local coordinate system, oriented to curve segment.

The main steps of the piecewise cubic fitting (PCF) process are shortly discussed below; the notation used in the discussion is introduced in Figure 6. To achieve a smooth curve, the curve fitting to any segment is constrained by its neighbors by enforcing an identical curvature at the segment connection points; in other words, PCF polynomial is continuous with its first derivative at connection points $x=s, x=t$, etc. The equations describing the third-order polynomial and its first derivative are:

$$S_k(x) = y_s + m_s \cdot (x-s) + a_s \cdot (x-s)^2 + b_s \cdot (x-s)^3$$

$$\text{slope} = S'_k(x) = m_s + 2 \cdot a_s \cdot (x-s) + 3 \cdot b_s \cdot (x-s)^2$$

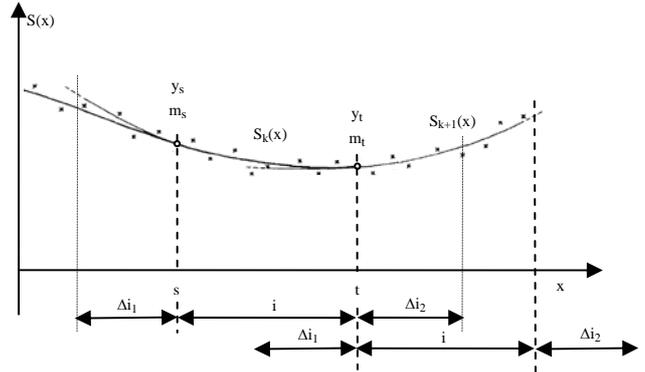


Figure 6. Piecewise weighted least squares curve fitting method.

The core processing includes the following steps: 1) a_s and b_s , the coefficients of the second and third order terms of the fitted curve for interval 'i' are estimated; consider the constant term (y_s) and the coefficient of the first order term (m_s) fixed, known from the curve fitting from the previous segment. In the adjustment, the points in interval $\Delta i_2 + i + \Delta i_1$ (past, present, and future data points) are used, 2) the value (y_t) and the slope (m_t) at $x=t$ are computed; these values as fixed constraints are used in the curve fitting for the next segment, and 3) step 1 is repeated to process the next segment.

Step 1	$y - y_s - m_s \cdot (x-s) = a_s \cdot (x-s)^2 + b_s \cdot (x-s)^3$ LS for points in interval $\Delta i_1 + i + \Delta i_2 \Rightarrow \hat{a}_s, \hat{b}_s$ $\Rightarrow S_k(x) = y_s + m_s \cdot (x-s) + \hat{a}_s \cdot (x-s)^2 + \hat{b}_s \cdot (x-s)^3$
Step 2	$\hat{y}_t = S_k(t) = y_s + m_s \cdot (t-s) + \hat{a}_s \cdot (t-s)^2 + \hat{b}_s \cdot (t-s)^3$ $\hat{m}_t = S'_k(t) = m_s + 2 \cdot \hat{a}_s \cdot (t-s) + 3 \cdot \hat{b}_s \cdot (t-s)^2$ $\rightarrow y_t = \hat{y}_t \text{ and } m_t = \hat{m}_t$
Step 3	$S_{k+1}(x) - y_t - m_t \cdot (x-t) = a_t \cdot (x-t)^2 + b_t \cdot (x-t)^3$ LS for points in interval $\Delta i_1 + i + \Delta i_2 \Rightarrow \hat{a}_t, \hat{b}_t$

4.2 Matching curves

Iterative registration algorithms are increasingly used for registering 2D/3D curves and range images recently. The well-known Iterative Closest Point (ICP) algorithm (Besl and McKay, 1992; Madhavan et al., 2005) is adopted here to match curves describing pavement markings obtained from LiDAR intensity and GPS measurements. The ICP algorithm finds the best correspondence between two curves (point sets) by iteratively determining the translations and rotations parameters of a 2D/3D rigid body transformation.

$$\min_{(R,T)} \sum_i \|M_i - (RD_i + T)\|^2$$

Where R is a 2×2 rotation matrix, T is a 2×1 translation vector and subscript i refer to the corresponding points of the sets M (model) and D (data). The ICP algorithm can be summarized as follows:

1. For each point in D , compute the closest point in M
2. Compute the incremental transformation (R, T)
3. Apply incremental transformation from step (2) to D
4. If relative changes in R and T are less than a given threshold, terminate, otherwise go to step (1)

Our 2D ICP was implemented in Matlab and space-scale optimization was incorporated to reduce execution time.

5. EXPERIMENTAL RESULTS

To perform an initial performance test of the proposed method, a typical intersection was selected from a recently flown LiDAR survey, where GPS-surveyed pavement markings were available. Figure 7 shows the area with linear pavement markings measured from the LiDAR intensity data as well as the GPS points. Note the clearly visible misfit between the two point sets; the horizontal accuracy of the GPS-surveyed points, provided by the Ohio Department of Transportation VRS system is 1-2 cm.

The LiDAR point-based description of the pavement markings was obtained by filtering. The search space was defined by the GPS control points (pavement markings are assumed to be within ±1 m of their true location) and intensity thresholding was used to extract the linear features; the road pavement has low intensity value while the pavement markings exhibit higher intensities. The threshold is adaptively defined by analyzing the histogram of the LiDAR intensity values of the road surface around the surveyed road pavement markings and/or by examining intensity values of road surface profiles (LiDAR scan-lines).

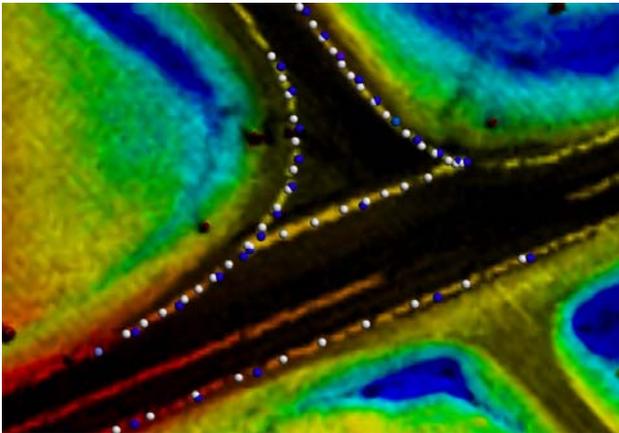


Figure 7. Intersection with pavements markings measured from LiDAR intensity data (white) and GPS-surveyed (blue).

In the curve-fitting step, both representations of the linear features are computed according the algorithm described in 4.1. Figure 8 shows one example of the fitted curves for the west curb line.

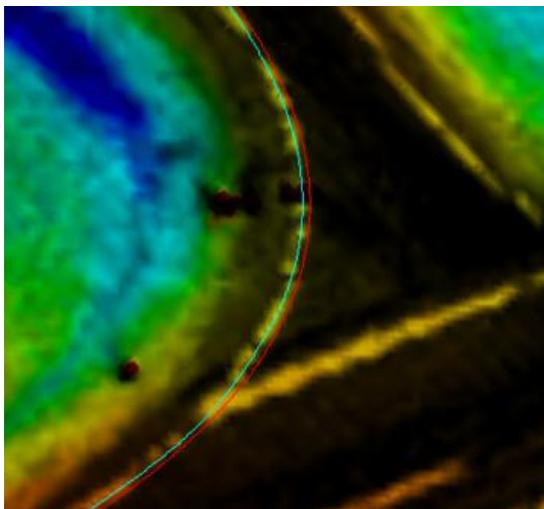


Figure 8. Curve fitting based on LiDAR and GPS points.

The results of the ICP-based curve matching for all the four curve lines is shown in Figure 9. Visually, the transformation shows a good fit; the blue points nicely fall on the GPS-defined curves. Note that the original curve points, derived from LiDAR, moved to the control curve similarly, as opposed to they would move if the individual curves had matched. Figure 10 shows the results of curve matching for the lower straight pavement line, including both the transformation results; as expected the individual transformation implements a perpendicular projection to the control curve.

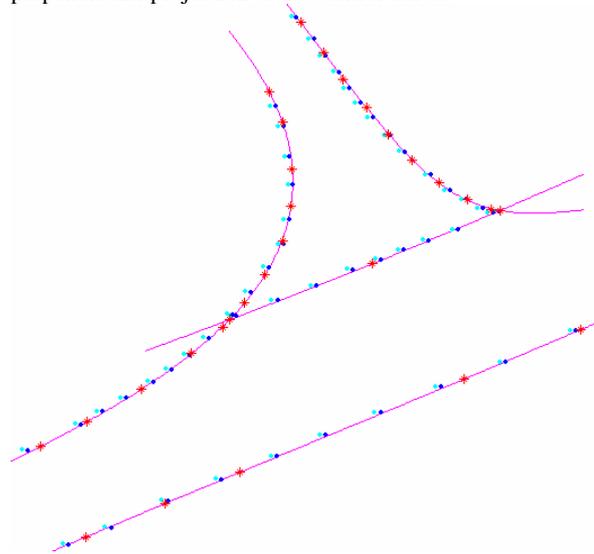


Figure 9. Curve matching based on all the four curves; magenta: curves fitted to control points, red: GPS control points, blue: curve points derived from LiDAR, and cyan: transformed curve points (derived from LiDAR).

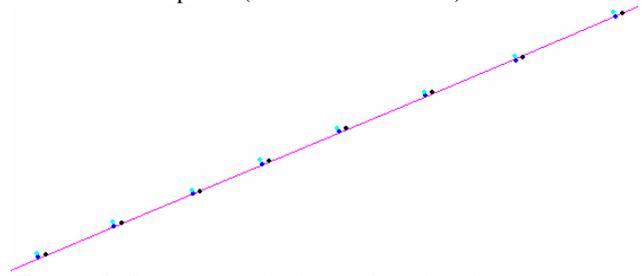


Figure 10. Comparing individual and combined curve fitting to a straight feature; magenta: reference curve, cyan: points derived from LiDAR, blue: transformed points based on single curve matching, and black: transformed points based on matching all the four curves together.

To assess in actual numbers the accuracy of the transformation, obtained by the ICP-based curve matching, correspondence between the LiDAR-derived curve and the control curve were established. Since the two curves in general are not totally identical, even after the final ICP iteration, the transformed LiDAR-derived points are close but not necessarily fall on the control curve. However, the location of the transformed LiDAR-derived points represents the best fit to the control curve in least squares sense. Therefore, these points are projected to the closest points of the control curve, and then they are considered as conjugate points. The transformation parameters between these two point sets (the original LiDAR-derived points and their corresponding points on the control curve) are calculated in a least squares adjustment. In this computation, the transformation parameters for the test data were estimated at $\sigma_{\Delta X} = \pm 0.013m$, $\sigma_{\Delta Y} = \pm 0.017m$, and $\sigma_{\text{angle}} = \pm 1.95 \text{ arcmin}$, indicating that a good match was found with the

ICP method for the spatially well distributed test data set. The numerical values, including the transformation parameters, error terms, and dispersion matrix are listed in Tables I and II.

Transformation parameter	ICP-adjusted results [m, °]	Estimated accuracy [cm, °]
ΔX	0.46	0.013
ΔY	-0.08	0.017
φ	-0.09	0.03

Table I. Transformation results (2D).

1.0e-003 *	0.1789	-0.1699	-0.0063
	-0.1699	0.2902	0.0087
	-0.0063	0.0087	0.0003

Table II. A posteriori dispersion matrix.

The ~2 cm horizontal accuracy is reasonable given the fact that the GPS-surveyed points are known at 1 cm-level accuracy and the LiDAR-based pavement marking positioning accuracy is estimated at the few cm range.

6. CONCLUSION

The introduced method to automate the use of pavement markings as ground control showed good initial performance. Both the curve fitting and ICP-based matching delivered robust results. Further research will consider the extension of technique to 3D.

References:

- Ahokas, E., Kaasalainen, S., Hyypä, J. and Suomalainen, J., 2006. Calibration of the Optech ALTM 3100 Laser Scanner Intensity Data Using Brightness Targets, Proceedings of ISPRS Commission I. Symposium.
- ASPRS LiDAR Committee, 2004. ASPRS Guidelines Vertical Accuracy Reporting for LiDAR Data http://www.asprs.org/society/committees/lidar/Downloads/Vertical_Accuracy_Reporting_for_Lidar_Data.pdf
- Baltsavias, E.P., 1999. Airborne Laser Scanning: Basic Relations and Formulas. ISPRS Journal of Photogrammetry & Remote Sensing, Vol. 54: 199-214.
- Besl, P. J. and McKay, N. D. A method for registration of 3-d shapes. IEEE Trans. Pat. Anal. and Mach. Intel. 14(2), pp 239-256, Feb 1992.
- Burman, H. .2002. Laser Strip Adjustment for Data Calibration and Verification. International Archives of Photogrammetry and Remote Sensing, 34 (Part 3A): 67-72.
- Crombaghs, M. J.E., R. Brügelmann, E.J. de Min, 2000. On the Adjustment of Overlapping Strips of Laseraltimeter Height Data. International Archives of Photogrammetry and Remote Sensing, 33, (Part B3/1):224-231.
- Csanyi N, Toth C., Grejner-Brzezinska D. and Ray J., 2005. Improving LiDAR data accuracy using LiDAR-specific ground targets, ASPRS Annual Conference, Baltimore, MD, March 7-11, CD-ROM.
- Csanyi, N. and Toth, C., 2007. Improvement of LiDAR Data Accuracy Using LiDAR-Specific Ground Targets, Photogrammetric Engineering & Remote Sensing, Vol. 73, No. 4, pp. 385-396.
- Filin, S., and Vosselman, G., 2004. Adjustment of Airborne Laser Altimetry Strips. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 34 (Part B) pp. 285-289.
- Hasegawa, H., 2006. Evaluations of LiDAR Reflectance Amplitude Sensitivity Towards Land Cover Conditions, Bulletin of the Geographical Survey Institute, Vol. 53.
- Ichida, K. and Kiyono, T.1977. Curve Fitting with One-Pass Method with a Piecewise Cubic Polynomial, ACM Transactions on Mathematical Software, Vol. 3, No. 2, pp. 164-174.
- Kager, H. and Kraus, K., 2001. Height Discrepancies between Overlapping Laser Scanner Strips. Proceedings of Optical 3D Measurement Techniques V, Vienna, Austria: 103-110.
- Kaasalainen, S., Ahokas, E., Hyypä, J. and Suomalainen, J., 2005. Study of Surface Brightness from Backscattered Laser Intensity: Calibration of Laser Data, IEEE Geoscience and Remote Sensing Letters, 2(3):255-259.
- Kilian J., Haala, N., Englich, M., 1996. Capture and Evaluation of Airborne Laser Scanner Data. International Archives of Photogrammetry and Remote Sensing, 31 (Part B3):383-388.
- Leica Geosystems, ALS50, <http://gis.leica-geosystems.com>
- Lutz, E., Geist, Th. and Stötter, J., 2003. Investigations of Airborne Laser Scanning Signal Intensity on Glacial Surfaces - Utilizing Comprehensive Laser Geometry Modeling and Orthophoto Surface Modeling (A Case Study: Svartisheibreen, Norway), proceedings of ISPRS Commission III, WG 3.
- Maas, H.-G., 2001. On the Use of Pulse Reflectance Data for Laserscanner Strip Adjustment. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 33 (Part 3/W4): 53-56.
- Madhavan, R., Hong, T., Messina, E. Temporal Range Registration for Unmanned Ground and Aerial Vehicles, Journal of Intelligent and Robotic Systems, Volume 44, Number 1 / September, 2005, pp. 47-69.
- Nobrega, R. and O'Hara, C., 2006. Segmentation and Object Extraction from Anisotropic Diffusion Filtered LiDAR Intensity Data.
- Optech, ALTM 3100AE, 2006, www.optech.ca/pdf/Brochures/ALTM3100EAWspecsfnl.pdf
- Schenk, T., 2001. Modeling and Analyzing Systematic Errors in Airborne Laser Scanners. Technical Notes in Photogrammetry, vol. 19. The Ohio State University, Columbus, USA.
- Sithole, G., 2005. Segmentation and Classification of Airborne Laser Scanner Data, Publication of Geodesy 59, Nederlandse Commissie voor Geodesie, Delft (184 pages).
- Song, J-H., Han, S-H., Yu, K, and Kim, Y., 2002. Assessing the Possibility of Land-Cover Classification Using LiDAR Intensity Data, International Archives of Photogrammetry, 34. pp. 4.
- Toth C., Csanyi N. and Grejner-Brzezinska D. 2002. Automating the Calibration of Airborne Multisensor Imaging Systems, Proc. ACSM-ASPRS Annual Conference, Washington, DC, April 19-26, CD ROM.
- Toth, C. and Grejner-Brzezinska, D., 2005. Geo-referenced Digital Data Acquisition and Processing Systems Using LiDAR Technology – Final report, ODOT State Job No. 147990.
- Vosselman, G., and Mass, H.-G., 2001. Adjustment and Filtering of Raw Laser Altimetry Data. Proc. OEEPE Workshop on Airborne Laserscanning and Interferometric SAR for Detailed Digital Elevation Models. OEEPE Publication 40, Stockholm, Sweden. pp. 62-72.
- Vosselman, G., 2002. On the Estimation of Planimetric Offsets in Laser Altimetry Data. International Archives of Photogrammetry and Remote Sensing, 34 (Part 3A): 375-380.
- Vosselman, G., 2002. Strip Offset Estimation Using Linear Features. 3rd International LIDAR Workshop, October 7-9, Columbus, <http://www.itc.nl/personal/vosselman/papers/vosselman2002.columbus.pdf>

REGISTRATION OF NEAR REAL-TIME SAR IMAGES BY IMAGE-TO-IMAGE MATCHING

B. Wessel^{a,*}, M. Huber^a, A. Roth^a

^a German Aerospace Center (DLR), 82234 Weßling, Germany - (birgit.wessel, martin.huber, achim.roth)@dlr.de

Commission IV, WG IV/3

KEY WORDS: Disaster, Mapping, SAR, Image, Registration, Orthorectification, Adjustment

ABSTRACT:

Near real-time SAR images are used at DLRs Center for Satellite Based Crisis Information for mapping natural hazards like flood disasters or landslides. In crisis situations a rapid and precise geocoding is mandatory to produce information within several hours that is comparable with other georeferenced information. Unlikely, the orbit information is often not very precisely known some hours after data acquisition, which leads to misregistration.

This paper deals with the development of a fully automatic orientation for SAR images. The goal is to determine a precise orbit for geocoding by image-to-image matching between a near real-time (NRT) image and a reference SAR image. The approach uses a new algorithm for feature extraction developed in computer vision by Lowe, 2004. No prior information about the pose of the images or the overlapping parts is needed. The point operator extracts points with scale- and rotation-invariant descriptors (SIFT-features). These descriptors allow a successful matching of image points even in situations with highly convergent images, i.e. with different incidence angles. Our approach consists of the following steps: Point extraction and matching by using the SIFT parameter descriptors with an extended matching scheme. The resultant points of the reference image are used as ground control points (GCPs) for an adjustment of the SAR imaging geometry of the NRT image. Then, a geocoding of the NRT SAR image can be carried out. This achieves results equivalent to a high precision orbit.

Examples of two datasets are presented and the results of the approach are discussed and evaluated. The results show that the approach can be used for a wide range of scenes with different SAR sensors, different incidence angles, and different overlap extensions. The results are very reliable but depend on well structured image content.

1. INTRODUCTION

One service of DLR is a rapid mapping service in case of larger natural disasters performed by the Center for Satellite Based Crisis Information. Near real-time satellite images are used to map natural hazards like flood disasters or landslides within several hours. As the interpretation often relies on actual and former information a precise geocoding of near real-time satellite data is a mandatory. But due to inaccurate orbit data this is often a difficult task. Inaccurate orbit data leads

- on the one hand to an incorrect geolocation that impede the joint interpretation with other data that often takes place in crisis context.
- On the other hand no correct orthorectification can be performed because the underlying DEM is misaligned to the SAR image. Hence, effects caused by the side looking SAR geometry (layover, foreshortening, shadow) cannot be corrected and so much the worse a misaligned DEM even introduces additional errors during the geocoding process to the SAR image.

The exterior orientation of a SAR image needed for geocoding is established by the sensor trajectory $S(X,Y,Z)$ over the time t , the Doppler frequency (if not zero), and the slant range R_0 to the first (near range) pixel. The frequency, time and range information are recorded by the sensor. The knowledge of a high precision orbit is often only several days to one month after the SAR image acquisition available. For recently acquired images mostly just predicted orbits are available.

Though, the predicted orbit of Envisat is within the order of few meters. Especially, RADARSAT orbits generally need substantial refinement as even the post-processing orbit can be shifted by several kilometres. As the overall philosophy in crisis mapping is to be able to process any image one can get, as fast and as good as possible the orbits have to be corrected.

So for near-real time applications some ground control points are needed to solve the sensor trajectory, if the preliminary orbit exceeds the desired accuracy. GCPs can be acquired from maps or other images. Crisis mapping mostly relies on former information, so image-to-image registration is an obvious method for registration. Mostly the GCP measurement is done manually for crisis application. This is a difficult task, as point identification and precise point determination in SAR images is often difficult.

1.1 Overview to related work

Registration of two SAR images is a standard task in InSAR Interferometric co-registration. Standard co-registration technique is based on cross-correlation measure, carried out on extended image patches. This method allows obtaining very high accuracies on image pairs characterized by sufficiently high correlation values. In areas with low coherence or different incidence angles correlation technique results are generally unsatisfactory. Although the normalized cross-correlation is invariant against mean gray-level it is not to local dissimilarities and rotation. To overcome the above limitations, a co-registration approach based on features has been indicated as a

* Corresponding author.

potential solution. The idea consists of exploiting the response of isolated and bright points to find the image local shifts by estimating their precise position in small patches. Cross-correlation over such small patches provides accurate position estimates (Adam et al., 2003).

The registration approaches for different SAR images like different viewing geometry or different sensors is a difficult topic and currently an ongoing research. An approach quite similar to ours is suggested by Auquière et al., 1998. They perform a simple transformation, i.e. a (2D-)shift, applied to the PRI images. Läbe and Förstner, 2006 used the SIFT operator (Lowe, 2004) for an automatic orientation of aerial images, which is even for optical images still a topic of research.

The development of a mostly automatic procedure for precise geocoding of near real-time satellite SAR images is on our focus in this paper. In the first part of the paper we explain an automatic approach for the generation of ground control points (GCPs) for precise orbit determination. It is based on image-to-image matching with a new developed feature extraction algorithm from computer vision (Lowe, 2004). After GCPs are found a least squares adjustment of the orbit parameters is performed. In the second part of the paper we present some results for this registration method and show in detail the found matches.

2. DESIGN OF A PROCEDURE FOR AUTOMATIC GEOCODING BASED ON IMAGE-TO-IMAGE REGISTRATION

We design a registration strategy based on image-to-image registration between a near real-time image and a former SAR image with a post-processing orbit to get precise orbit information. The registration strategy is as follows:

- Getting an archive SAR product with a high-precision orbit, preferred in slant range, e.g. Envisat ASAR
- Getting a near real-time satellite image, e.g. from RADARSAT or Envisat ASAR.
- Extraction of sub-pixel precise interest points in both slant range images with Lowe's SIFT keypoint operator.
- Matching of the keypoints from both images.
- Calculation of the exact 3-D position of the matched keypoints from the high-precision orbit image and a DEM to get GCP's.
- Least squares adjustment of the imprecise orbit by means of the GCPs, including blunder elimination of incorrect matches.
- Final geocoding of the near real-time image on the basis of the refined orbit parameters.

Figure 1. gives an overview of the procedure. In the following the individual steps are discussed in detail.

2.1 Input data

Main inputs for the procedure are one geocoded reference image and one near real-time sensed image to be rectified. Here, we focus on registering SAR images. The matching can be performed on geocoded as well as on slant range images. To assure a maximum of similarity, the matching takes place on two slant range images to neglect possible errors. Additionally, required for geocoding is a digital elevation model (DEM) and orbital information. The resolution of the DEM should be in the same order as the aimed image resolution. The orbit

information should be high-precision for the reference image and should be known by approximation for the sensed image.

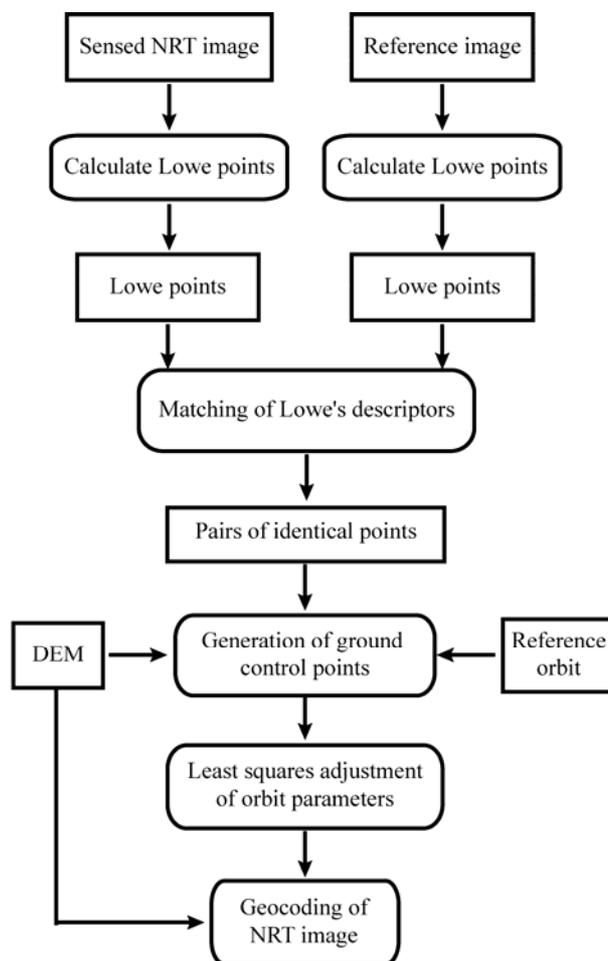


Figure 1. Procedure for automatic geocoding with image-to-image registration. The individual steps are discussed in Sect. 2.

2.2 Feature extraction

Generally, significant features are extracted from images to perform a feature matching. Most sophisticated point operators from optical images like the Harris or Förstner operator are not that successful on SAR data, because of the speckle effect. Therefore, for tie-point extraction we chose a newly developed keypoint extractor from Lowe, which was developed for optical images but because of its multi-scale approach it seems especially suitable for SAR data. The computational efficiency and robustness support our task of a fast near real-time geocoding in the same way.

The Lowe keypoint detector (Lowe, 2004) is based on a multi-scale approach and finds scale-independent point features, which are additionally extremely invariant to rotation and affine transformations. Therefore, it can cope with different incidence angles. Samples of found features are presented in Figure 4 and Figure 8.

2.3 Feature matching

For matching of two extracted point data sets no further assumptions about the orientation are made. The matching

relies on a vectorial description for each point (Lowe, 2004) and is performed by calculating the Euclidian distance between the vectors. The shorter the distance is, the more likely a homologous point pair is present. The use of that kind of vectorial point description instead of image patches and image correlation allows a very computational efficient and fast matching.

Some additional restrictions are introduced, because incorrect orbit parameters cause mainly shifts in and across the flight direction. So the search space was restricted to a certain distance and to a certain connection angle between a homologous point pair.

2.4 Generation of GCPs

The matched feature points have to be transformed into control information by the geometry of the reference image. Input to this transformation is the pixel coordinates resp. slant range coordinates of the homologous points of the reference image, for which typically a high-precision orbit is available. The individual points of the reference image can be geocoded with the help of a DEM and we get 3D-ground control coordinates (= object coordinates) of the corresponding points. These can be used as ground control points for adjusting the orbit parameters.

2.5 Least squares adjustment of orbit parameters

The next step is an adjustment for determining the near real-time orbit with the help of GCPs. A classical least squares adjustment is performed to estimate the unknown orbit parameters.

The position of the sensor depends on the time. It can be approximated locally with high accuracy by an orbit polynomial, which is derived from annotated sensor positions (state vectors). The functional model is described in the following.

Functional model for orbit determination: The equations used for the adjustment are the same used for geocoding. To be demanded are the parameters of the orbit polynomial that is approximated by a polynomial of forth degree

$$S(i) = \vec{a}_0 + \vec{a}_1i + \vec{a}_2i^2 + \vec{a}_2i^3 + \vec{a}_2i^4 \tag{1}$$

with $S = [X_S, Y_S, Z_S] =$ coordinates of the sensor
 $i =$ azimuth/time

The Doppler and range equations give the relation between ground and image coordinates in a SAR image (see Figure 2 or Roth et al., 1993):

$$F_1(i, j) = f_{DC} - \frac{2(\vec{P} - \vec{S})(\vec{P}' - \vec{S}')}{\lambda |\vec{P} - \vec{S}|} \tag{2}$$

$$\vec{R} = ||\vec{P} - \vec{S}||$$

$$F_2(i, j) = R_0 + m_r j - |\vec{P} - \vec{S}| \tag{3}$$

- where $i, j =$ image coordinates (i : azimuth/time; j : range)
- $\vec{P} = [X, Y, Z] =$ coordinates of mapped pixels on earth ellipsoid
- $R =$ slant range vector between sensor S and earth location P
- $\vec{P}', \vec{S}' =$ velocity of P resp. S
- $\lambda =$ radar wavelength
- $f_{DC} =$ Doppler frequency shift
- $R_0 =$ slant range to first image row
- $m_r =$ pixel spacing

As the orbit is only approximately known for real-time applications the adjustment solves iteratively the radar equations (Eqs. 2 and 3) by adjusting the orbit polynomial until the range and Doppler equations are simultaneously fulfilled. The adjustment contains a blunder elimination that eliminates those GCPs from the calculation that derivation is significant greater than the standard deviation of the iteration result. When the iteration result is found the adjusted orbit can be used for geocoding.

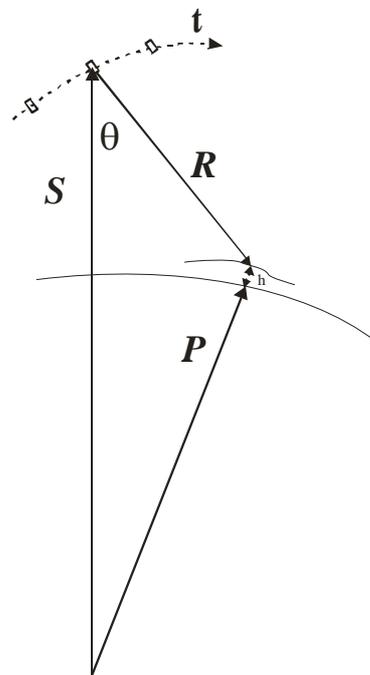


Figure 2. Functional sensor model of the SAR acquisition.

2.6 Geocoding of near-real time image

The last step of the procedure is to geocode the sensed NRT SAR image with the help of the adjusted orbit parameters. For geocoding the indirect rectification method is used. For each output pixel, which defines a co-ordinate triple (easting, northing, height) in the output map projection, the corresponding azimuth and range positions in the input image are determined. The Equations (1) to (3) are used for this step. The geocoding can be performed by means of interpolative or rigorous approach, respectively (Huber et al., 2004, Roth et al., 1994).

Note, that with the described method for image-to-image registration no previously known GCPs or manual measured GCPs are necessary.

3. EXAMPLES FOR NEAR REAL-TIME GEOCODING

The procedure has been applied to various SAR data sets as well as to aerial optical images. The tests presented here stem from a high-resolution aerial SAR data set and from a real crisis data set in Lebanon. The tests are carried out with a demonstration version of the Lowe keypoint extractor (<http://www.cs.ubc.ca/~lowe/keypoints/>). This version is restricted to images with a maximum image height or width of 1800 pixel. Therefore, we reduced the SAR images by a factor of eight instead of using the original image sizes. But as the points are extracted sub-pixel precise the loss of accuracy was smaller than expected. Also image reduction corresponds to a filtering and therefore reduces the speckle.

3.1 Matching of multi-temporal images

The first test site is based on two E-SAR images. The images were acquired in August 2002 and March 2003 and show the DLR in Oberpfaffenhofen, Germany. Both are acquired with almost the same incidence angles and overlap 100 %, though the reference image covers a larger scene. For the near real-time approach it is assumed that a high-precision orbit is only available for the first image. The extraction of Lowe points is done in slant range geometry. The resolution is reduced by a factor of eight that corresponds to a multilooking of eight pixels and a pixel spacing of about 8 m.

Figure 3 shows a closer view on some matched keypoints to demonstrate which kind of points are extracted and matched. The found corresponding points are mainly brighter or darker than its surroundings. Points 1 and 2 in Figure 3 are at the border of a darker to a brighter image region. Point 3 in Figure 3 is a round dark feature. As the extractor works in multiple scales all points are singular at a specific scale. One can note that by the bright keypoint 4, obviously it has significance in a smaller image scale.

The total number of extracted keypoints is above 1.000 for each image. The number of matched points is 75. Remarkable is the reliability of the matches. Nearly all found matches are correct. Only one mis-match is present (Figure 5). In Figure 5 the 75 matched keypoints are overlaid on the two images. Each keypoint pair is connected by a line. Correct matches are characterized by parallel lines. The amount of 75 was sufficient to solve the Eqs 1, 2, and 3 and to adjust the orbit of the second E-SAR image. The result of geocoding with the adjusted orbit is shown in Figure 4. Its accuracy is within half a pixel (0.5 m). Though, the keypoint extraction took place on 8 m pixel. This emphasizes the subpixel capability of the Lowe operator.

3.2 Matching of images with different incidence angle and from different sensors

The second test site is located in Lebanon and stem from a real crisis charter call to detect the oil spill in August 2006. The terrain is hilly to mountainous (see Figure 6). For this test site ENVISAT and RADARSAT images are available. This test site is more difficult than the previous since the images

- are from different sensors,
- have different incidence angles (22° - ENVISAT resp. 45° - RADARSAT), and
- the overlap is about 50% but without the water areas just 25 %.

Nevertheless, 6 matches were found. Figure 8 shows in detail the found matching partners between the Envisat and the Radarsat image.

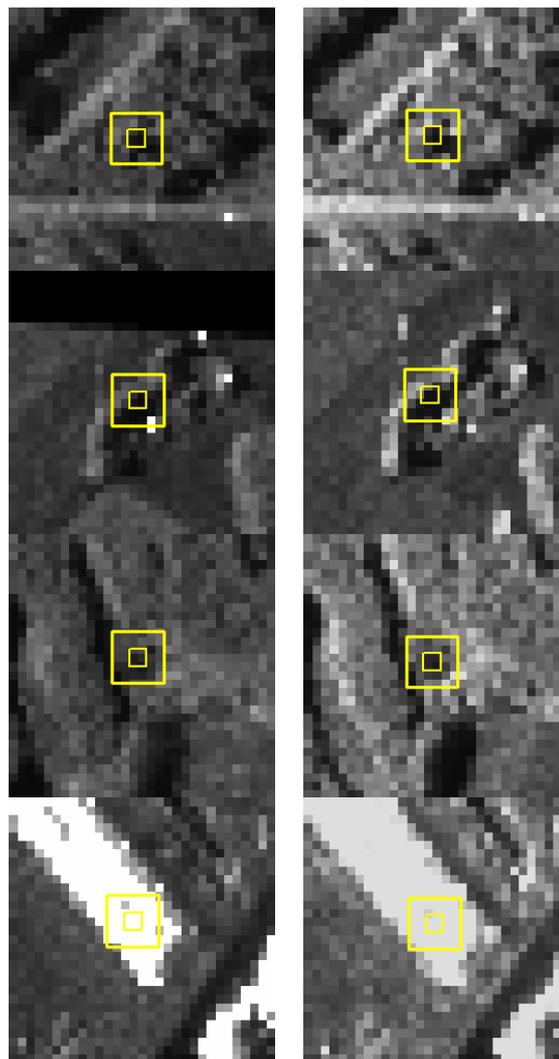


Figure 3. Left the new “near real-time” E-SAR image right reference image

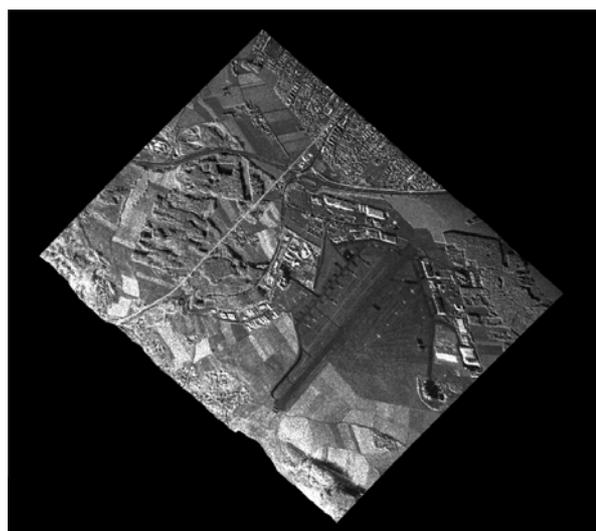


Figure 4. Geocoded E-SAR image by an adjusted orbit based on image to image registration

Though, the dimensions of the found features are different, the detection of features darker or brighter than its surrounding is equal to the behaviour of the high-resolution keypoints.

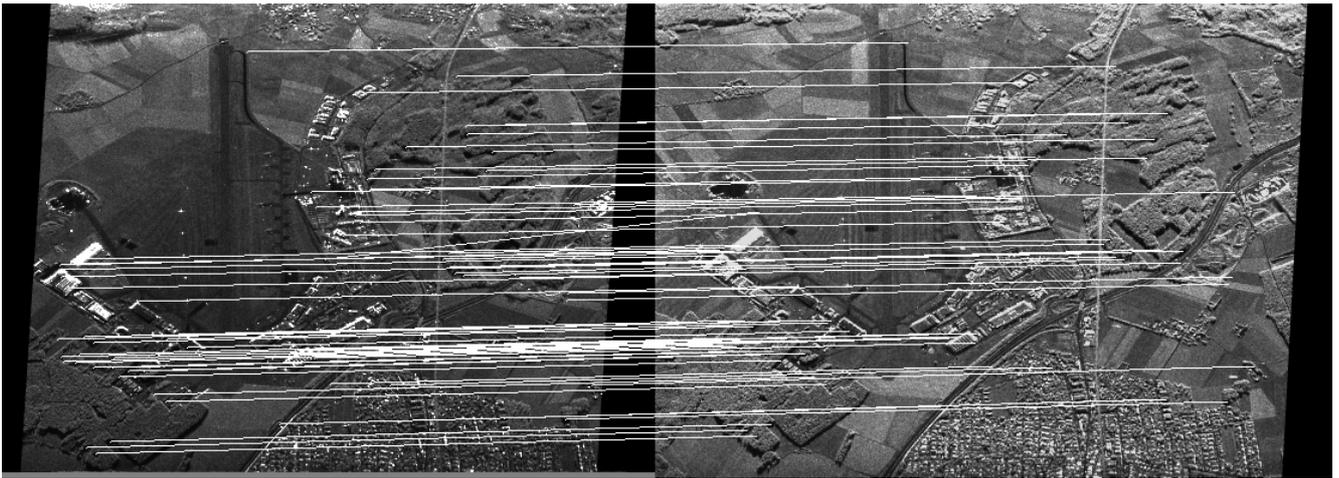


Figure 5. Results of matched SIFT keypoints on two E-SAR scenes in slant range geometry. The white lines connect two matched keypoints with each other: on the left the new “near real-time” E-SAR image on the right the reference image. 75 Matches were found.

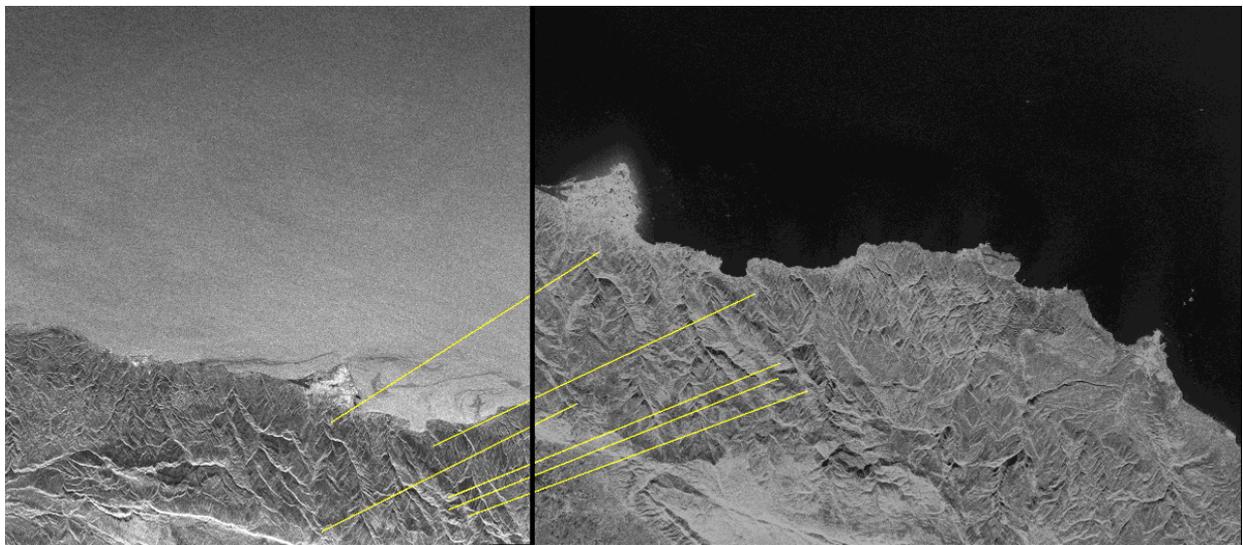


Figure 6. Lebanon test site results of matched SIFT keypoints between an ENVISAT scene (left) and a RADARSAT scenes (right) in slant range geometry. The yellow lines connect two matched keypoints with each other. 6 Matches were found.

But for reliable orbit determination a larger amount of GCP's would be necessary. At least more than 9 points are needed for orbit determination. As this wasn't possible in the test scene, no adjustment of SAR geometry was possible.

But such a limited number of points still can be used for a coarse image registration. The concept of a two-stage registration is fairly spread in optical image registration. At a reduced resolution some good points for a coarse registration are extracted and e.g. the coefficients of an affine transformation are calculated. Then, a fine matching takes place. Such a registration strategy, i.e. a two-stage registration, can be highly recommended for matching of SAR data. First of all, this technique is successfully used in optical data registration procedures. Secondly, it has the great advantage for SAR images that in very coarse resolution the speckle almost cancels out. And thirdly, for SAR data does a large amount of fine registration techniques from InSAR exist, which necessarily need good coarse registration.



Figure 7. Digital elevation model of the Lebanon test site.

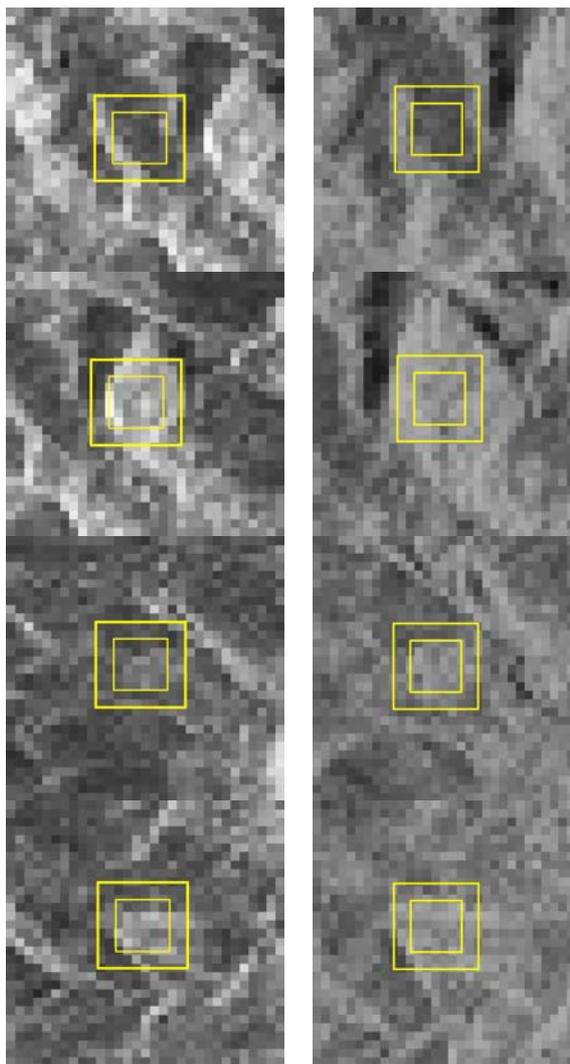


Figure 8. Examples of four extracted and matched SIFT image pairs from RADARSAT (left) versus ENVISAT (right)

4. SUMMARY AND OUTLOOK

The goal of the presented procedure aims to an automatic geocoding of near real-time images for crisis applications. The registration strategy relies on image-to-image matching of SAR images on the basis of the newly developed Lowe feature extractor. The results presented show that the Lowe operator performs surprisingly well on SAR images. In high resolution E-SAR images with forested and urban structures sufficient keypoints could be extracted to perform an adjustment of the orbit geometry for geocoding. The matches are characterized by very high reliability with a small rate of mismatches. Also on strongly different images with different incidence angles, different orbits and different sensors (ENVISAT versus RADARSAT) some reliable matches could be found in mountainous areas. Although the number was too low for orbit determination, the results are promising since the Lowe keypoint extraction is very fast and robust. One drawback seems that the keypoint operator strongly depends on well structured image content.

Further investigations on a refined strategy for keypoint extraction into a separated coarse and fine registration like it is used for optical images will be carried out. Also a matching

with the DEM will be investigated. For low structured scenes additional features like lines or contours should be extracted to improve the registration.

REFERENCES

- Adam, N., Kampes, B., Eineder, M., Worawattanamateekul J., and Kircher, M. 2003. The development of a scientific permanent scatterer system. In: *Proceedings of the ISPRS/EARSel Workshop "High Resolution Mapping from Space"*, Hannover (Germany)
- Auquière, E., Hau, P & Defourny, P. (1998). A fast track geocoding method for high temporal resolution SAR time series. In: *Proceedings of the Workshop on "Retrieval of Bio- and Geo-Physical Parameters from SAR Data for Land Applications"*, 21.-23. October 1998, ESTEC, Noordwijk, The Netherlands
- Huber M., Hummelbrunner W., Raggam H., Small D., Kosmann D. (2004), Technical Aspects of Envisat-ASAR Geocoding Capability at DLR. In: *Proceedings of the ENVISAT/ERS Symposium*, Salzburg, Austria, 6-10 September 2004.
- Läbe; T., Förstner, W, 2006. Automatic relative orientation of images. In: *Proceedings of the 5th Turkish-German Joint Geodetic Days*, March 29th – 31st, 2006, Berlin.
- Lowe, D., 2004. Distinctive image features from scale-invariant keypoints. In: *International Journal of Computer Vision*, 60(2), pp. 91-110.
- Lowe, D. 2005. SIFT demo program, Version 4, July 2005, <http://www.cs.ubc.ca/~lowe/keypoints/> (accessed 11 Jul. 2007)
- Roth, A., Craubner, A, Hügel, T., 1993. Standard geocoded ellipsoid corrected images. In: Schreier G. (Ed.): *SAR Geocoding and Systems*, Wichmann, Karlsruhe, pp. 159-172.
- Roth A., Huber M., Kosmann D. (2004), Geocoding of TerraSAR-X Data. In: *Proceedings of the 20th International Congress of the ISPRS*, Istanbul, Turkey, 12-23 July 2004, Commission III, pp. 840-844.

A MULTI-LEVEL IMAGE DESCRIPTION MODEL SCHEME BASED ON DIGITAL TOPOLOGY

Guobin Zhu ^{a,*}, Xiaoli Liu ^b, Zhige Jia ^c Qingquan Li ^b

^a International School of Software, Wuhan University, 129 Luoyu Road, Wuhan 430079, China -zhuu_gb@china.com

^b State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan 430079, China - liuxl.j@163.com

^c Institute of Seismology, China Seismological Bureau, Xiaohongshan Road, Wuhan 430071, China - jia.zg@126.com

Commission III, WG III/4

KEY WORDS: Digital Image, Image engineering, Image Segmentation, Multi-level Model, Digital Topology, Object Extraction

ABSTRACT:

It's necessary to discuss the topology model of digital images for integrating remote sensing and geographic information system in higher levels. Based on complex theories, a multi-level hierarchical image representation is presented that preserves topological relation equivalency and a set of functional architectures that efficiently reflects this representation. In the proposed hierarchical framework, a novel progressive region growing method is proposed that incorporates spatial information related to adjacency between pixels. The particularity of this method is that connected regions and their topology generate objects in different scales, furthermore constructing a tree-object structure reflecting their spatial relationships.

1. INTRODUCTION

Image engineering consists of three layers: low-level image processing based on pixel, inter-level image analyzing based on object, and high-level image understanding based on semantics (Zhu, 2003). The course from image processing, analyzing to understanding is a progressive procedure reflecting information processing phases. In the processing, images are interpreted by description models using certain ways; while the transition from digital number within low-level origin images, objects within inter-level image extraction to knowledge retrieved from high-level architecture (e.g. spatial relationship) is progressively accomplished. It is amazed that such a hierarchical image information representation that preserves spatial relationships is constructed along with the progressive processing, which implies an information framework from abstract status to physical one and a hierarchical transition from discrete structure to continuous one. However, problems remain along with the evolvement. Digital images are discrete objects in nature, but they are usually representing continuous objects or, at least, they are perceived as continuous objects by visual perception of human beings. Therefore, having a discrete and continuous representation of one object at the same time may activate interests.

The field of digital topology grew out of this challenge, and its main purpose is to study the topology properties of digital images (Rosenfeld, 1979). Based on the theory, this paper presents a multi-level hierarchical image representation that preserves topological relation equivalency between discrete and continuous descriptions of the same object and a set of functional architectures that efficiently reflects this representation. This hierarchical framework involves a general method to associate each digital object, in an arbitrary digital space, with a Euclidean polyhedron (named as its continuous

analogue), which naturally represents the "continuous perception" that an observer may take on that object. The multilevel architecture and, particularly, continuous analogues of objects can be applied to obtain new results in digital topology, by translating the corresponding continuous results through the levels of the architecture. So it may derive interests for integrating remote sensing and geographic information system in higher levels.

Image segmentation aim to subdivide the image into disjoint subsets of pixels, called regions, on the basis of some homogeneity criterion. Usually, this is the starting point in practical applications like content-based image retrieval or image compression (Chamorro-Martinez et al., 2003). Many types of segmentation techniques have been proposed in literatures (Chamorro-Martinez et al., 2003; Smith et al., 2000; Metternicht, 2003). Nevertheless, a drawback of most of these approaches is that they don't take into account that a region must be topologically connected. As a consequence, pixels belonging to separate and different regions could be assigned to the same cluster. The proposed hierarchical framework in this paper preserves spatial relationships and, so raises a suitable condition for image segmentation to incorporate spatial information related to adjacency between pixels. Based on the hierarchical representation, a novel progressive region growing method is proposed which incorporates spatial information related to adjacency between pixels. The particularity of this method is that connected regions and their topology generate objects in different scales, furthermore constructing a tree-object structure reflecting their spatial relationships.

* Corresponding author: Tel.: +86-27-68778916; fax: +86-27-68778910.
E-mail address: zhuu_gb@china.com (G. Zhu)

2. BASIC NOTIONS OF DIGITAL IMAGE

A topological model is used to explicitly specify adjacency and inclusion relationships between the different cells (vertices, edges, faces for dimensions 0, 1, 2) of a geometrical object. Information, called embeddings (labels), such as geometrical ones (for instance vertex coordinates) can be added to the model. The archetypical device model (\mathbb{R}^2) is a standard decomposition of Euclidean plane (\mathbb{R}^2) into abstract cells. The kernel of our modeller is based on the topology of cellular complexes which is the only possible topology of finite sets (Kovalevsky, 1989; Kovalevsky, 2006). Under this topology no contradictions or paradoxes arise when defining connected subsets and their boundaries. In this section, we briefly summarize the basic notions of the multilevel architecture based on the topology of cellular complexes, as well as the notation that will be used throughout the paper.

Definition 1: An abstract cellular complex (AC complex) $C=(E, B, dim)$ is a set E of abstract elements (cells) provided with a natural reflexive, antisymmetric and transitive binary relation $B \subset E \times E$ called the bounding relation, and with a dimension function $dim: E \rightarrow I$ from E into the set I of non-negative integers such that $dim(e') \leq dim(e'')$ for all pairs $(e', e'') \in B$. An AC complex C , when regarded in this way as a topological space, is called an underlying polyhedron and written $|C| = \cup \{e, e \in C\}$. Moreover, C satisfies the following properties:

- C1: \leq is a partial order in E .
- C2: If $e' \in C$ and e'' is a face of e' then $e'' \in C$.
- C3: If $e', e'' \in C$ then $e' \cap e''$ is a face of both e' and e'' .
- C4: $\forall e \in C$ is a face of some n -dimensional cells in the C .

According to the definition of open star in the general topology, we introduce two types of "digital neighbourhoods": star and extended star of a cell $x \in C$ in a given digital object $O \subseteq cell_n(C)$.

Definition 2: The star of x in O is the set $st_n(x; O) = \{y \in O, x \leq y\}$ of n -cells (pixels) in O having x as a face. Similarly, the extended star of x in O is the set $st_n^*(x; O) = \{y \in O, x \cap y \neq \emptyset\}$ of n -cells (pixels) in O intersecting x .

Definition 3: Given an AC C and two cells $x, y \in C$. A centroid-map on C is a map $\theta: C \rightarrow |C|$ such that $\theta(x)$ belongs to the interior of x ; that is, $\theta(x) \in x^0 = x - \partial x$, where $\partial x = \cup \{y, y < x\}$ stands for the boundary of x .

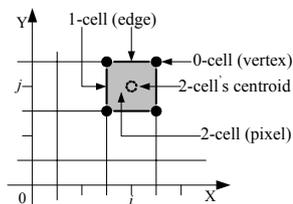


Fig. 1 An abstract cell in \mathbb{R}^2

The maximum dimension of the cells of an AC complex is called its dimension. We shall mainly consider complexes of dimension 2. Their elements with dimension 0 (0-cells) are called points, elements of dimension 1 (1-cells) are called edges, elements of dimension 2 (2-cells) are called pixels (faces). As a consequence, the spatial layout of pixels in a

digital image is represented by a device model, which is an n -dimensional locally finite AC complex determined by the collection of unit 2-squares in \mathbb{R}^2 whose edges are parallel to the coordinate axes and whose centres are in the set \mathbb{Z}^2 . The centroid-map we will consider in \mathbb{R}^2 associates to each square x its barycentre $\theta(x)$. In particular, if $dim(x)=2$ then $\theta(x) \in \mathbb{Z}^2$. So that, every digital object O in \mathbb{R}^2 can be identified with a subset of points in \mathbb{Z}^2 .

Let the digital image (G) be a locally finite AC complex of dimension n ($n=2$). Each n -cell in G represents a pixel, and so the digital object ($O \in G$) displayed in a digital image is a subset of the set $cell_n(G)$ of n -cells in G , denoted by $O \subseteq cell_n(G)$; while the other lower dimensional cells in G are used to describe how the pixels could be linked to each other. If the full image is partition into m disjoint subsets ($O_i, i=1, 2, \dots, m$), then

$$\bigcup_{i=1}^m O_i = G \text{ and } \forall i \neq j : O_i \cap O_j = \emptyset \quad (1).$$

The 2-cells (pixels) are the area elements. In image processing, 2-cells must be associated with the notion of pixels since a gray value assigned to a pixel originates from measuring the amount of energy radiated from an elementary area. Valid adjacencies are between adjacent pixels which are labelled by identical image values (Rosenfeld, 1984). In fact, pixels are usually a combination of materials, and frequently in multispectral and hyperspectral image (Bragato, 2004; Plaza et al., 2004). To solve this problem, region is regarded as a fuzzy subset of pixels (Bragato, 2004), in such a way that every pixel of the image has a membership degree to that region. It is a key that a fuzzy resemblance relation between neighbour pixels is obtained from a fuzzy resemblance relation between their corresponding feature vectors. So we characterize each pixel p by means of a vector of features $\vec{f}_p = [f_p^1, f_p^2, \dots, f_p^k]$, where a feature $f_p^i \in \mathbb{R}$ with $i \in \{1, 2, \dots, k\}$, is a numerical measure of any relevant characteristic that may be obtained for p . In general, we define a fuzzy resemblance relation between feature vectors as a fuzzy subset FR of $\mathbb{R}^k \times \mathbb{R}^k$, with membership function $FR: \mathbb{R}^k \times \mathbb{R}^k \rightarrow [0, 1]$. For simplicity, we define a set of centroids and compute the membership value for all the pixels in the image to each centroid.

3. MULTI-LEVEL IMAGE REPRESENTATION

3.1 A Hierarchical Information Representation (HIR) for Images

In the processing from low-level to high-level, images are interpreted by description models using certain ways in different abstract levels; while the transition from digital number within low-level origin images, objects within inter-level image extraction to knowledge retrieved from high-level architecture is progressively accomplished. Amazing as it is, such a hierarchical information representation (HIR) for images that preserves spatial relationships is constructed along with the progressive processing, which shows an information framework from abstract to material and a hierarchical transition from discrete to continuous. Considering the gap between the discrete world of digital objects and the Euclidean world of their continuous interpretations, this paper proposes a progressively hierarchical image representation, in which there

coexist discrete and continuous descriptions of the same object preserved topological relation equivalency.

This framework consists of four levels: two extreme levels: device and continuous levels, and additional two intermediate levels: logical and conceptual levels. The device level represents the physical problem (Fig.2a/a') whereas the models in the continuous level are topological spaces which allow us to use the well-known results of continuous topology (e.g. polyhedral topology) (Fig.2b/b'). The additional two intermediate levels are used to bridge the two extreme levels, which allow a progressive evolution from the discrete object to the Euclidean one, and vice versa. The logical level is closer to the device level and it is used for processing, for writing algorithms and to show their correctness (Fig.2c/c'). The conceptual level is the nearest to the continuous level and it is used to translate results and notions from the continuous level to the logical level (Fig.2d/d').

3.2 A Multi-Level Functional Architecture

Because a digital space fixes, among all the possible continuous interpretations, just one for each digital object, this continuous interpretation of a digital object is represented at each level of the architecture using a different model; in particular, the corresponding model at the continuous level is a Euclidean polyhedron, called the continuous analogue of the object. According to target applications, in each one of them we can use different models.

To illustrate our purpose, we introduce an empirical functional structure in dimension 2 that efficiently reflects the proposed representation. Inside the modelling structure, objects are represented with four different models: discrete, discrete contours, discrete analytical, and continuous. In the hierarchical structure, links between the different consecutive levels allow us to manipulate and propagate modifications locally. Updating the whole structure is thus not systematically needed. Of course such a framework comes with a prize. The complexity of the hierarchical structure is much more complex than classical topology based modelling or imaging softwares.

Let the full image G be subdivided into n disjoint objects $O_i \subseteq cell_2(G)$ ($i=1,2,\dots,n$), which is a set of 2-cells with domain f_p^* of homogeneous feature value (Fig. 2).

3.2.1 0-Device Level: corresponds to the classical discrete representation of image elements (pixels) in a computer screen. Each pixel is represented by a square topological face associated with a colour feature embedding. Moreover, integer coordinates are attached to each topological vertex.

A discrete model of O in this level is the subcomplex $D_O^f = \{x \in G; x \leq y, y \in O\}$ induced by the cells in O , and $\vec{f}_p \in f_p^*$ for any cells (Fig. 2a). This level has a very few degree of abstraction and we only represent the physical aspects of the objects.

3.2.2 1-logical level: corresponds to the contours obtained for each 4-connected region with homogeneous feature vector. The representation of these contours is based on the inter-pixel model (Kovalevsky, 1989). Each discrete point is represented by a 0-cell and two successive points are linked by a 1-cell (Fig. 2b). This representation simplifies the coverage of level 0 regions boundaries. There is no embedding at this level and geometrical information needed for the visualization of the level are located in level 0 and can be accessed from level 1.

A discrete border model of O in this level is an undirected graph L_O^f , whose vertices are 0-cells and linked by 1-cells in O , moreover two of those cells x, y are adjacent if there exists a common face $a \subseteq x \cap y$ such that $\vec{f}(x), \vec{f}(y) \in f_p^*$.

In this level, we consider the proximity aspects of the objects and so, we can study some properties of topological nature. The main function of this level is to be the support for writing the algorithms and to prove their correctness. Because L_O^f is not planar, this level is far from the mathematical model. So we need the conceptual level as an interface between the level above and the continuous level.

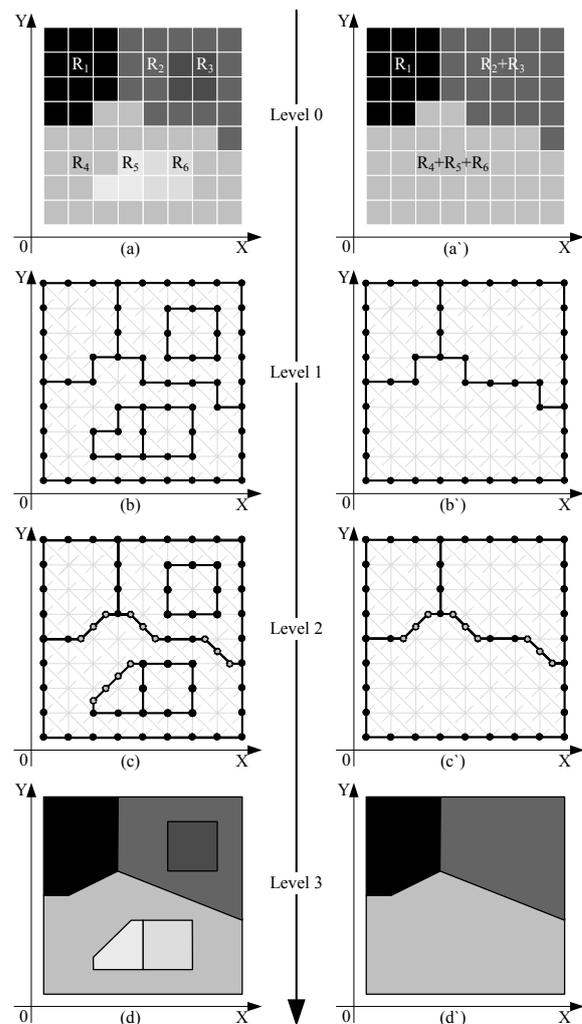


Fig. 2 Multi-level hierarchical representation for digital images: (a) discrete level; (b) logical level; (c) conceptual level; (d) continuous level.

3.2.3 2-Conceptual Level: is an implicit representation of the discrete border primitive. It corresponds to the discrete analytical description of the level 1 region contours. More precisely, each contour is described as a discrete analytical polygon computed according to the given models. For the sake of simplified solution, we flatten out the graph L_o^f in a natural way and we get the planar graph (Fig. 2c). In this graph there are two different kinds of vertices: 0-cells and centroids of 1-cells. Observe that this graph is a triangulation of the Euclidean plane, which makes up the conceptual level.

A discrete analytical model of O is an induced directed graph C_o^f , whose vertices are the centroids $\Theta(x)$ of all 0-cells and 1-cells $x \in G$ with $\vec{f}(x) \in f_p^*$, and its directed edges are pairs $(\Theta(x), \Theta(y))$ with $x < y$.

3.2.4 3-Continuous Level: is an explicit representation. In this level, each region is described as a Euclidean polygon with attribute features in the classical boundary representation form. The primitive of this level may be created using the tools available inside the modeller, or may be the result of the reconstruction process applied on level 0. 2D Euclidean vertex coordinates and face features are associated to the continuous model of O in this level (Fig. 2d).

A simplicial analogue S_o^f is an order complex associated to the directed graph C_o^f . This is, $\langle \Theta(\alpha_0), \Theta(\alpha_1), \dots, \Theta(\alpha_m) \rangle$ is m -simplex of S_o^f if $\Theta(\alpha_0), \Theta(\alpha_1), \dots, \Theta(\alpha_m)$ is a directed path in C_o^f . This simplicial complex defines the simplicial level for the object O in the architecture and, finally, the continuous analogue of O is the underlying polyhedron $|S_o^f|$ of S_o^f .

When given a concrete problem, we must choose specific models in each level and functions which can support the functionality that we have described. Specifically, suppose that these chosen models are D, L, C and S for the device, logical, conceptual and continuous level, respectively. Let $\Omega(D), \Omega(L), \Omega(C)$ and $\Omega(S)$ be the sets of the objects of these models. Then here are several mappings: (1) a 1-1 mapping $i: \Omega(D) \rightarrow \Omega(L)$; (2) a natural mapping $\pi: \Omega(L) \rightarrow \Omega(C)$ and $\pi^*: \Omega(C) \rightarrow \Omega(L)$; (3) a suitable mapping $j: \Omega(C) \rightarrow \Omega(S)$. So we have the following functional architecture:

$$\begin{array}{ccc} \Omega(L) & \xrightarrow{i} & \Omega(D) \\ \pi \downarrow \uparrow \pi^* & & \\ \Omega(C) & \xrightarrow{j} & \Omega(S) \end{array}$$

This architecture provides a link between the discrete world of digital pictures represented by a cellular complex, and a Euclidean space through several other intermediate levels and, embodies the transitions from low-level feature to high-level semantic. Further, this framework involves a general method to associate each digital object, in an arbitrary digital space, with a Euclidean polyhedron called its continuous analogue, which naturally represents the "continuous perception" that an observer may take on that object. The multilevel architecture and, particularly, continuous analogues of objects can be applied to obtain new results in digital topology, by translating the corresponding continuous results through the levels of the architecture. Thus it may be interesting for integrating

geographic information system and remote sensing in higher levels.

4. A ROOT TREE STRUCTURE OF OBJECTS

The proposed hierarchical framework preserves spatial relationships, raising a suitable condition for image segmentation to incorporate spatial information related to adjacency between pixels. Thus this multi-level image representation allows manipulating and propagating modifications locally (Fig.2, Fig.3). We locally modify the segmentation of the digital image by filling the fragmental regions in order to obtaining desired-only regions. Thus, we can determine which cells must be removed in the other levels. Indeed, in the original digital image, these cells were surrounded by two faces are represented with different colors. After the edits, these cells are surrounded by two faces with the same color. Using the links between level 0 and level 1, we can easily find the cells of level 1 having been removed. And so on for level 2 and level 3. Here, the main interest of using links of the structure is that we can make local modification without recomputing the entire structure.

Suppose the full image and the discrete level correspond to two extreme levels of objects: the root and leafs of a tree, respectively. Starting with an arbitrary pixels, objects and features can be extracted easily through a down-top region merging cluster in different scales. While performing the down-top union of regions (children) at one level into a single larger region (parent) at the next higher level of the tree, regions are grouped together according to similarities between their feature vectors, which include such features as colour information, orientation, texture, size, energy, and neighbour information. Any other levels of segmented regions lie betwixt the two extreme levels. Hence a root tree structure is constructed in a simple and "natural" way the regions according to their topology and, where each level consists of two data structures, a weight graph $W_{level}(N_{level}, E_{level})$ and a disjoint-region set R_{level} (Corme, 1990).

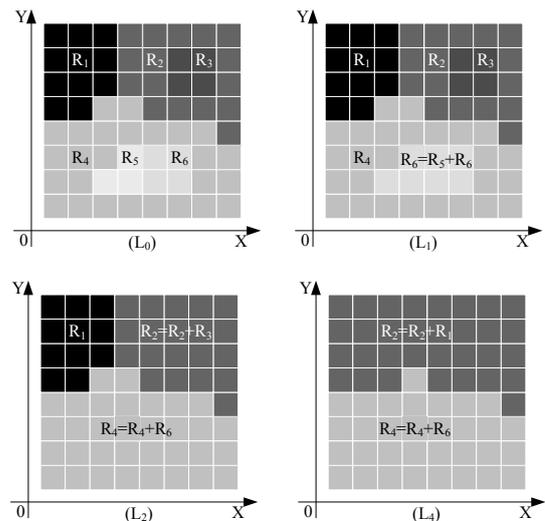


Fig. 3 Discrete segmentation operation based on the multi-level image representation

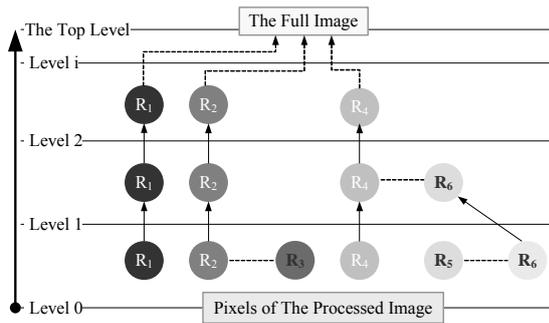


Fig.4. A root tree structure of objects

In the tree construction (Fig. 4), the root node denotes the border of the whole image where all the regions are included, the leaf nodes denote all pixels of the discrete level, other nodes of the tree denote regions N_{level} and has a feature vector characterizing those regions, each region in R_{level} is one node or a group of nodes from the next lower level. Two nodes are linked by an edge if and only if their corresponding regions are adjacent each other. The weighted graph in each level is composed of the nodes and edges within this level. The undirected dot-edges in each level define the spatial relationships between the nodes, and the weight of each edge defines the degree of the similarity between those two nodes. The relationship among the disjoint-set in each level is illustrated by the groupings. The directed solid-edges of regions in different levels reflect the inclusion relationships between them. Since regarded as a rooted “directed” tree, and in which from a given region stored, there is only one path to the root, this is, the map from the parent node to child nodes is $1: n (n \in Z)$. Given this directed notion of a rooted tree, a rooted sub-tree can be defined for each node of the tree. As a consequence, for each region which contains included regions, a sub-tree is recursively generated. The parent node shall directly inherit attribute features of the first child with smallest feature distance among all the edges connected to this node.

The tree-like organization embodies reflecting the relationships of regions and their attribute features. For each region stored in the tree, we can directly access to the included regions, the parent region, (e.g. vertical navigation), or adjacent regions (horizontal navigation) (Maire et al., 2005). Once the tree structure is stored, further complementary manipulating is possible for local modification without recomputing the entire structure.

5. DISCRETE IMAGE SEGEMENTAION

As a practical instance, a region growing algorithm that efficiently constructs this root tree structure is proposed. The sequential process is described as four steps.

5.1.1 Step 1. Initialization: To reduce the computation complexity, the image is initially partitioned into small $K \times K$ (K varies by image size, usually equates to 4 or 8) blocks composed of topological connected pixels with similar image features. Each blob has an associated set of features measured from the original pixel spectral features of the image. Based on these extracted features and the neighboring relationships among blobs, the one-level is built.

5.1.2 Step 2. Region Merging Cluster: A recursive node clustering and region merging are performed at each level using a bottom-up strategy. At the end of each iteration, the algorithm has completed one level of the hierarchy, so a new level is constructed and the structure is updated. The procedure is repeated until the stopping condition has been attained, which is defined as either the desired final number of objects, or the maximum feature distance (threshold) below which clusters may be combined. Or if not specified, the algorithm will continue until a full tree structure of the original image is built with the root node of the tree being a single object corresponding to the whole image.

5.1.3 Step 3. Extraction of Geometry and the Adjacency Relation: Geometry and the topological relation in each level are extracted for each region detected during step 2, and stored in the database. When progressively recognizing, classifying and integrating image objects (nodes) from different regions in intra-level or inter-level, besides just colour and geometry information, multi-dimensional information, including the orientation, texture, size, energy, and neighbour relationships between objects are considered in the processing of region merging cluster.

5.1.4 Step 4. Building the tree structure: Based on segmentation results, and by analyzing topology relations and feature vectors, the regions are recursively created and stored in the tree structure and, so region inclusions are recursively deduced and propagated to the whole segmented image by adjacencies relations. Hence the tree-like organization of the hierarchical relations of spatial objects reflects hierarchical topological relations of spatial objects. From this tree structure, objects and features can be extracted easily through a top-down traversal of the final hierarchical structure of the image.

In our practical instance, to ensure low computation complexity, the second low computation complexity, the step 2 requires that each single-node region merge with at least one other region in this level. So, before a new level is generated, this iteration guarantees that all sets of the current S_{level} include at least two nodes. As a consequence, each new level of the hierarchy is guaranteed to have no more than half the number of nodes as the previous level, ensuring fast convergence of the algorithm.

6. CONCLUSION AND FUTURE WORK

In this paper we presented a multi-level hierarchical information representation (HIR) for images that preserves topological relation equivalency and a set functional architecture that efficiently reflects this representation. Each level corresponds to a particular representation of the same object: discrete, discrete border, discrete analytical of regions, and continuous representations. Each level is linked with the levels above and below itself by transition mapping. This ensures the topological relation equivalency between all the representations. In the proposed hierarchical framework, a progressive region growing method is used to subdivide image into regions and to construct a tree-object structure reflecting their spatial relationship. The particularity of this method is that it incorporates spatial information related to adjacency between pixels, while keeping connecting regions and their topology generated in different scales. As a short term goal, we plan to develop spatial relationship of discrete digital objects, while more studies in details for the propagation of local

modifications along with the hierarchical multi-level will be pursued in the future.

7. REFERENCES

Zhu, G. 2003. Remote Sensing Image Analysis Based on Hierarchical Multi-resolution Structures. *Geomatics and Information Science of Wuhan University*, 28:315-320.

Rosenfeld, A., 1979. Digital Topology. *Am Math Month*, 86:621-630.

Chamorro-Martinez, J., Sanchez, D., 2003. A Fuzzy Color Image Segmentation Applied to Robot Vision. In *Advances in Soft Computing - Engineering, Design and Manufacturing*, J. M. Benitez, O. Cordon, F. Hoffman, and R. Roy, eds. (Springer), pp. 129-138.

Smith, J. R., Li, C.-S., 2000. Image Classification and Querying Using Composite Region Templates, Vol 77, Elsevier Science.

Metternicht, G. I., 2003. Categorical fuzziness: a comparison between crisp and fuzzy class boundary modelling for mapping salt-affected soils using Landsat TM data and a classification based on anion ratios, Vol 168, Elsevier Science.

Kovalevsky, V., 1989. Finite Topology as Applied to Image Analysis. *Computer Vision, Graphics, and Image Processing*, 46:141-161.

Kovalevsky, V., 2006. Axiomatic Digital Topology. *J Math Imaging*, 26:41-58.

Rosenfeld, A., 1984. The fuzzy geometry of image subsets. *Pattern Recognition Letters*, 2:311-317.

Bragato, G., 2004. Fuzzy continuous classification and spatial interpolation in conventional soil survey for soil mapping of the lower Piave plain, Vol 118, Elsevier Science.

Plaza, A., Martinez, P., 2004. A new approach to mixed pixel classification of hyperspectral imagery based on extended morphological profiles, Vol 37, Elsevier Science.

Cormen, T. H., Leiserson, C. E., and Rivest, R.L., 1990. *Introduction to Algorithms*, MIT Press, Cambridge, MA.

Maire, C., Datcu, M., 2005. Object and topology extraction from remote sensing images. *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, Vol.2, Iss., pp: II-193-6.

Keyword Index (ISPRS)

Accuracy	107	Laser Scanning	1, 7, 13
Adjustment	161, 179	LIDAR	25, 107, 139, 167, 173
Aerial	47, 83	Mapping	7, 47, 113, 179
Algorithms	31	Matching	25, 89
Analysis	19, 65, 107, 139	Measurement	53
Automation	31, 101	Method	53
Building	13, 65, 77, 95, 101	Mobile	7
Calibration	151, 167	Model	7, 47, 65, 145, 167, 185
Camera	83, 151, 161	Modelling	41, 71, 77, 95
Change	71	Multiresolution	89
Change Detection	145	Multisensor	107
City	7	Multispectral	25, 89
Classification	1, 25, 77	Object	19, 25, 185
Close Range	7, 53, 59, 65	Optical	71, 125
Data	25, 77, 145	Orientation	161
Database	131	Orthorectification	95, 179
DEM/DTM	25, 71, 77, 101, 113	Parameters	83, 151
Detection	1, 83, 89, 119, 125	Processing	25
Digital	47, 53, 95, 145, 185	Quickbird	89
Digitisation	31	Radar	157
Disaster	179	Raster	31
Edge	101	Reconstruction	13, 53, 59
Environment	95	Registration	1, 179
Error	71, 107	SAR	19, 35, 113, 179
Estimation	151	Satellite	125, 157
Exterior	161	Scale	47
Extraction	25, 31, 65, 77, 101, 173, 185	Scene	19
Feature	31, 173	Segmentation	1, 31, 139, 185
Forestry	139	Semi-Automation	31
Fusion	7, 25, 145	Sensor	71, 125
Fuzzy Logic	119	Sequences	65
GPS/INS	107	Structure	151
High Resolution	131, 157, 161	Surface	47, 145
IKONOS	119	Systems	101, 107
Image	25, 31, 47, 65, 77, 83, 95, 131, 179, 185	Terrestrial	7, 65
Imagery	13, 125, 157	Test	59
Information	25	Texture	65
Infrared	65	Thermal	65
Integration	13	Three-Dimensional	13, 41, 65
Interpretation	19	Tracking	41
Knowledge Base	101	Understanding	95
Land Cover	101	Urban	65, 77, 89, 95
Large	47	Value-Added	157
		Vector	31, 131

Keyword Index (Complementary)

Airborne	107	Multi-View	167
Bayesian Generative Model Based Stereo ...	167	Parameter Estimation	151
Bursting Test	59	Particle Swarm Optimization	119
Class Mean	119	Phase Unwrapping	35
Enhanced Image Products	158	Position	151
Error Maps	71	QA/QC	173
Error Propagation	107	Road	83, 119
Error Sources	107	Road Junction	131
ETFE-Foils	59	Self-Calibration	167
Evaluation	167	Series	83
Eye-Tracking	41	Shape	59
Fast Moving Target	125	Spaceborne	158
Geoinformation Products	158	Stereo	71, 77, 167
Geospatial	131	Strain and Stress Behaviour	59
Gradient Vector Flow	131	TerraSAR-X	158
Homography	151	Three Line Camera	161
HRSC	161	Topology	185
Human-Computer Symbiosis	41	Traffic	83
InSAR	113	UltraCamX	47
Intensity	173	Uncertainties	71
Interferometry	35	Vegetation	139
Interior Orientation	161	Vehicle	83
Knowledge Elicitation	41	Velocity	151
Level of Detail	13	VHR	77, 125
Millimeterwaves	35	X-Band	158
Motion	151	Ziplock Snake	131
Multi-Level	185		

Author Index

Argialas, D. P.	101	Heurich, M.	139
Bachleitner, M.	83	Hinz, S.	59, 89
Baltsavias, E.	25	Hoegner, L.	65
Barr, S. L.	13	Höfler, H.	7
Baulig, C.	7	Huber, M.	179
Beinat, A.	1	Jalobeanu, A.	71
Blug, A.	7	Jia, Z.	185
Boehmsdorff, S.	35	Kiraci, A. C.	31
Brehm, T.	35	Kirchhof, M.	151
Brzezinska, D.	173	Krauß, T.	77
Charmette, B.	83	Krzystek, P.	139
Chikomo, F. O.	13	Kurz, F.	83
Crosilla, F.	1	Leitloff, J.	89
Dambacher, M.	7	Leonhardt, A.	83
Dell'Acqua, F.	19	Levin, E.	41
Demir, N.	25	Li, Q.	185
De Rouck, J.	53	Lisini, G.	19
De Wulf, A.	53	Liu, X.	185
Eker, O.	31	Marinov, B. D.	95
Essen, H.	35	Mavrantza, O. D.	101
Fitzenz, D. D.	71	May, N. C.	107
Gamba, P.	19	Mercer, B.	113
Gienko, G.	41	Mills, J. P.	13
Goossens, R.	53	Mohammadzadeh, A.	119
Gruber, A.	47	Pagot, E.	125
Gutjahr, K.	125, 157	Pakzad, K.	131
Hanssens, P.	53	Paska, E.	173
Heipke, C.	131	Pesaresi, M.	125
Hennau, M.	53	Raggam, J.	157

Ravanbakhsh, M.	131	Stilla, U.	35, 65, 77, 89, 139
Reinartz, P.	77, 83	Strecha, C.	167
Reitberger, J.	139	Suri, S.	83
Rist, F.	59	Tavakoli, A.	119
Rosenbaum, D.	83	Thoennesen, U.	167
Roth, A.	179	Toth, C.	107, 173
Rottensteiner, F.	145	Valadan Zoej, M. J.	119
Schäfer, R.	151	Van Damme, D.	53
Scherer, N.	151	Van Damme, L.	53
Schiemann, L.	59	Van Gool, L.	167
Schmidt, N.	157	Visintini, D.	1
Schneider, S.	47	von Hansen, W.	167
Şeker, D. Z.	31	Weber, M.	157
Sepic, F.	1	Wessel, B.	179
Spangler, M.	83	Wimmer, A.	157
Spiegel, M.	161	Wölfelschneider, H.	7
Stätter, R.	83	Zhu, G.	185
Stephani, M.	59		