AUTOMATED TREE AND BUILDING DETECTION IN RESIDENTIAL AREAS FROM STEREO IKONOS IMAGES

J.R. Kim [a, *], J.-P., Muller [a]

[a] Dept. of Geomatic Engineering, University College London, Gower Street, London, WC1E 6BT UK
*jkim@ge.ucl.ac.uk

WG IV/3

KEY WORDS: Stereo IKONOS, Tree and building detection, DTM, image matcher

ABSTRACT:

Stereoscopic analysis of target images is the primary and conventional method for photogrammetric 3D range data construction. For object detection in high resolution satellite images, it is advantageous if the processing method can exploit a stereo derived DEM so that the technique can be applied anywhere in the world. Central to this is the need for a very reliable stereo matching algorithm and an improved camera model to perform such an analysis. Object detection using a stereo DEM is described in this article, including optical image processing and data fusion techniques and how these were applied to the IKONOS image.
Results are here presented for stereo IKONOS images over several residential areas which are occupied by dense woodland and small housing building unis at a rural location called Barton-Bendish in central England. Derived bare earth DTMs from DEM show < 3m rms height difference compared with OS® PANORAMA. Also, a quality assessment by visual inspection of object detection results shows reasonable accuracy. The tree detection process gives a higher detection ratio than 85% consistently and the building detection process attains up to a 80% detection ratio even with relatively small scale targets.

1. INTRODUCTION

High quality 3D range data produced through direct range measurement by an active sensor (usually LiDAR) is a very useful tool for automated landscape object detection. However, such data is not usually available and is very expensive to collect. In such cases, spaceborne stereo very high resolution imagery can be considered as a replacement. Stereo DEMs from high resolution stereo satellite images or aerial photos can have a higher resolution than direct range measurement products such as LiDAR but the vertical accuracy of individual height points is not comparable and is usually lower for spaceborne imagery. Therefore, automated boundary detection of landscape objects in a stereo DEM is usually not very reliable. The purpose of this research is the development of new stereo matching schemes for high-resolution satellite stereo image of residential areas, specifically IKONOS, to address such object boundary delineation problems and apply them to the automated object detection of trees and buildings when combined with multi-spectral clues.

To start with, camera information from the IKONOS images (Rational Polynomial Coefficient (RPC)) is updated using bias correction with a few GCPs and it will be demonstrated that a maximum 3m positional accuracy can be obtained.

To build a more reliable DEM, epi-polarity is exploited and combined with the Adaptive Least Squares Correlation (ALSC) method. We developed an optimization scheme for epi-polar constraint – based on a cooperative algorithm by Zitnick and Kanade (2000).

The object detection process consists of a combined processing chain consisting of first a bare Earth construction (DTM) routine from the stereo height points, followed by a multi-spectral classification scheme employing normalized height points as training vectors. It should be noted that this approach has already successfully been applied to a LiDAR based object detection (Muller et al., 2004). The DTM is applied to derive normalised height points which are feed-forwarded to a local classification stage to split building segments, tree and grass areas. Then through a channel point detection and ellipse fitting, individual tree crowns are split and building outlines are simplified to polygons.

2. BACKGROUND

2.1 Data sets description

Data over the Barton-Bendish test area in Eastern England comprises a stereo pair of IKONOS panchromatic 1m and 4m multi-spectral images covering an 11 x 11km area. Table 1 and Figure 1 show the main image characteristics and the resized outlines of the IKONOS images. As described in the Open GIS specifications (2001) on RPC, the original image pair is divided into two pieces to maintain the accuracy of the RPC coefficients and their epi-polarity.

Processing Level	Standard Geometrically Corrected
Image Type	PAN/MSI
Interpolation	Bicubic
Stereo	Stereo
Datum	WGS84
Nominal Collection Azimuth	Right :344.3 deg, Left :233.7 deg
Nominal Collection Elevation	Right : 69.4 deg, left 70.9 deg

Table 1. IKONOS image characteristics for the Barton Bendish test area
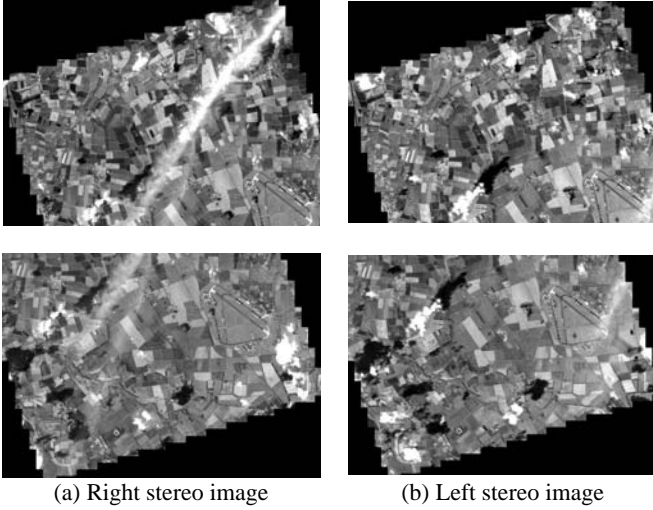
| (a) Right stereo image | (b) Left stereo image |

Figure 1. IKONOS Stereo image pair over Barton Bendish area split into 2 sections as described in the text

2.2 Updating of IKONOS sensor models

IKONOS Geo products normally have 15-metre CE90 positioning accuracy (Space Imaging, 2005). It means that the sensor model needs to be updated to derive accurate 3D data. RPC (Rational Polynomial Coefficient) is provided instead of physical sensor information with each IKONOS image. Therefore, the first starting point in IKONOS stereo processing is to update the RPCs. A detailed specification for RPCs is fully described in the Earth Image Geometry models of the Open GIS abstract specification (Open GIS Consortium, 2001). Basically it uses a ratio of two polynomial functions to transform ground X, Y and Z coordinates to image row and column as shown as:

$$r_1 = \frac{p_1}{q_1} \qquad (1)$$

$$c_1 = \frac{p_2}{q_2}$$

where p_1, p_2 are the numerators of the RPCs and q_1, q_2 are the denominators of the RPC.

Usually the format of a generic polynomial can be given as:

$$p = \sum_{i-0}^{m1} \sum_{j=0}^{m2} \sum_{k=0}^{m3} a_{ijk} X n^i Y n^j Z n^k \qquad (2)$$

where a_{ijk} is the polynomial coefficient in i,j,k order, X_n is the x ground coordinate , Y_n is the y ground coordinate and Z_n is the elevation of the ground level above a reference surface.

Usually with commercial optical images, a 3rd order polynomial results in 80 coefficients (20 for each p_1, q_1, p_2, q_2) - the maximum power of ground coordinates is limited to 3 such that the polynomial coefficient should be set to zero if $i+j+k$ is more than 3.

A simpler method was employed based on the results of analysing positional error properties of IKONOS. Here, the RPC update is tried with only 2 constant terms. As seen in the previous section, most of the error is directional such that by moving the space plane in the (x,y) direction, which is labelled

as a bias, an improvement in accuracy is expected of at least several pixels.

Hanley et al. (2002) proposed a bias corrected RPC using the following relationships:

$$r + A_o + A_1 r + A_2 c = \frac{p_1(X,Y,Z)}{p_2(X,Y,Z)} \qquad (3)$$

$$c + B_o + B_1 r + A_2 c = \frac{p_3(X,Y,Z)}{p_4(X,Y,Z)}$$

where $A_o, ... B_2$ are the bias factors, (r,c) are the image row and column and $p1, .. p4$ are the RPCs.

Therefore bias-corrected RPCs, incorporating shift terms A_0, B_0 can be given by:

$$p_1(X,Y,Z) = (a_1 - b_1 A_o) + (a_2 - b_2 A_o)X + ... + (a_{20} - b_{20}A_o)Z^3 \qquad (4)$$

$$p_3(X,Y,Z) = (c_1 - d_1 A_o) + (c_2 - d_2 B_o)X + ... + (c_{20} - c_{20}A_o)Z^3$$

We tested the accuracy of bias-corrected RPCs with GPS and paper map based GCPs and check points (see Table 2).

For each GCP, the back-projected row and column value can be calculated from an updated RPC. On the other hand, real row and column values can be identified in the image plane.

Number of GCPs	Number of Check points	RMS error of check points		Std.dev of checks shift	
		x	y	x	y
27	23	2.275	2.557	2.752	3.970
35	15	2.002	2.229	3.192	2.823

Table 2. Errors of bias compensated RPC

Another approach is to update IKONOS positional information by using an 3D affine transformation, which is even simpler than RPC (Fraser et al., 2001a and 2001b) as follows:

$$x = A_1 X + A_2 Y + A_3 Z + A_4 \qquad (5)$$

$$y = A_5 X + A_6 Y + A_7 Z + A_8$$

The basic assumption of their research is that high correlations exist between exterior orientation parameters with a perspective projection. The narrow bundle of rays effectively approaches a skew parallel projection as the field of view angle of the IKONOS sensor is less than 1 degree. Using a multi-image orientation/triangulation, sub-pixel accuracy can also be achieved. Table 3 shows the positioning accuracy by affine transformation.

GCP number	Check point number	RMSE error of check points		Standard deviation of check points	
		x	Y	X	y
26	25	3.639	3.404	3.150	4.188
34	17	2.503	2.028	3.137	2.258

Table 3. Positional accuracy of IKONOS affine camera model

Both methods show high reliability but in this research, Bias-corrected RPCs were employed as the default. In some areas

where RPC intersection did not converge, affine transformation was used instead of RPC updates.

3. ALGORITHMS

The processing flow is similar to the methods described by Muller et al. (2004). The primary difference is the source of 3D measurements. Hence the focus was on the stereo matching process with subsequent steps such as DTM construction and height-multi-spectral data fusion were somewhat modified considering the characteristics of 3D stereo measurements.

3.1 Image matching

Urban areas incorporate a number of discontinuities. Therefore, a specially optimized image matcher is necessary because the traditional Sum of Squared Differences (SSD) or Normalised Cross Correlation (NCC) cannot address problems associated with discontinuities in the image. The UCL in-house Pyramidal Gruen-Otto-Chau (P-Gotcha) (Day et al., 1992) used Gruen's Adaptive Least Squared Correlation (ALSC) (Gruen, 1985) for the cost calculation and aggregation of costs. Previously, Otto and Chau (1989) developed a region-growing method as the global optimizer of disparity, but the method was not suitable for areas containing any significant discontinuities.

This study uses ALSC for cost calculation and aggregation but a different method for global optimisation. The method uses a strategy for global optimisation that involves introducing a pre-processing matcher prior to ALSC. This pre-processing matcher (hereafter referred to as the "up-front matcher" or UFM) was designed to exploit epi-polarity and suitably address discontinuities.

The UFM consists of two main components: a simple corner point-matcher based on a Log-Polar transformation to define the disparity range and the cooperative algorithm from Zitnick and Kanade (2000). The "costing" part of the ALSC was modified by the use of Okutomi and Kanade (1992)'s adaptive window method.

The overall structure of the image matcher for 3D range data extraction for urban areas is shown in Figure 2 and the details of the individual components will be elaborated in a future peer review publication.
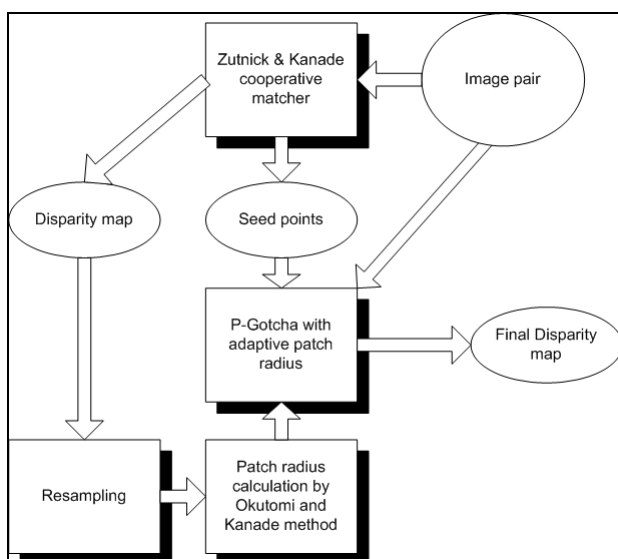

Figure 2. The overall structure of the image matcher for stereo IKONOS

This image matching flow appears to address the delineation problem in the landscape object boundary as seen in the examples in Figure 3.
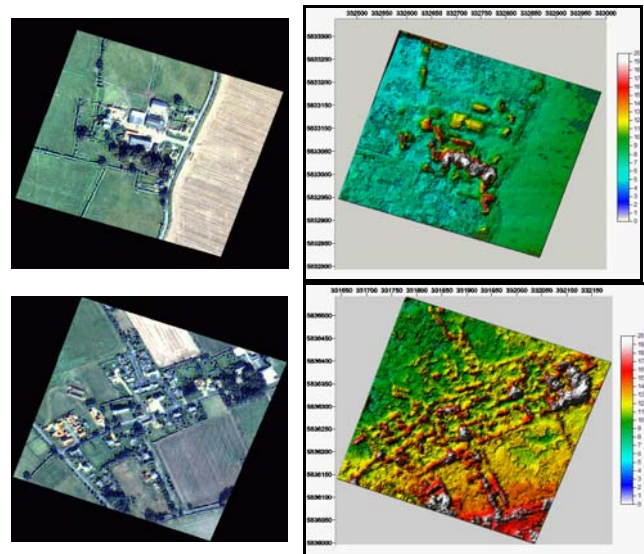

Figure 3. Examples of stereo DEMs using the modified image matching scheme

3.2 DTM construction

The construction of a DTM was an indispensable precursor process for object detection. A DTM construction algorithm was used which was originally developed for Lidar data using local extrema detection, hierarchical refinement using height residual, and gridding. The method described by Chaudhuri and Shankar (1989) is employed for local extrema detection. Then the slope analysis and region growing updated 'natural surface' and Smith and Wessel's (1990) minimum curvature method is used as the gridding component. A newly constructed "bare earth surface" is initially subtracted from the original DEM heights and an iterative process is applied to refine the mask of natural surface. Further details on these algorithms are given in Muller et al. (2004). Here we used stricter slope criteria to split "natural" and landscape object surface considering the characteristics of the stereo DEM. This "bare earth" DTM is integrated with NDVI to produce a preliminary building and tree targets through thresholding. These potential building and tree targets are known as Region of Interests (RoIs).

3.3 Object detection

3.3.1 Buildings: Building footprint refinement and generalisation occur in three consecutive stages of multi-spectral processing. In the first stage, supervised classification is applied using the training vectors defined by NDVI together with 3D information (i.e. the RoIs). A 3D point distribution check is applied twice to fuse the results of the supervised classification with the segments produced by Fuzzy Clustering and Fuzzy Merging (FCFM) (Looney, 2002) of the multi-spectral image. Then a seeded region growing (SRG) method (Adams and Leanne, 1994) is applied to extract the correct building roof parts. An algorithm for construction 3D building models is not developed here; instead, a simple boundary generalisation is performed. Further details can also be found in Muller et al. (2004).

3.3.2   Trees: This stage consists of two steps: supervised classification to discriminate grass and trees and individual tree crown fitting and splitting. The latter comprises iterative morphological filtering and ellipse fitting (Fitzgibbon et al., 1996). The grass and tree splitting scheme is almost identical with the classification stage of building detection. With NDVI and n-DEM (formalised heights of the landscape objects above the DTM) values, a supervised classification was applied to high NDVI area. Then the channel points between individual tree crowns' pixel intensity are detected by the method of Wood (1996). By applying morphological erosion continuously around the channel points, the tree crown can be segregated as seen in Figure 4 (a). The remnant of these iterative operations is called the "core" of the tree crown (Figure 4 (b)) and an ellipse can be fitted on the reconstructed boundary of the core. In our method, the eccentricity of fitted ellipses can be used as a verification parameter for the detected trees. If the fitted ellipses do not satisfy the criteria, then the failed part is used to provide feedback to unprocessed "preliminary" tree parts. The final result shows reasonable reliability as can be observed in Figure 4 (c).
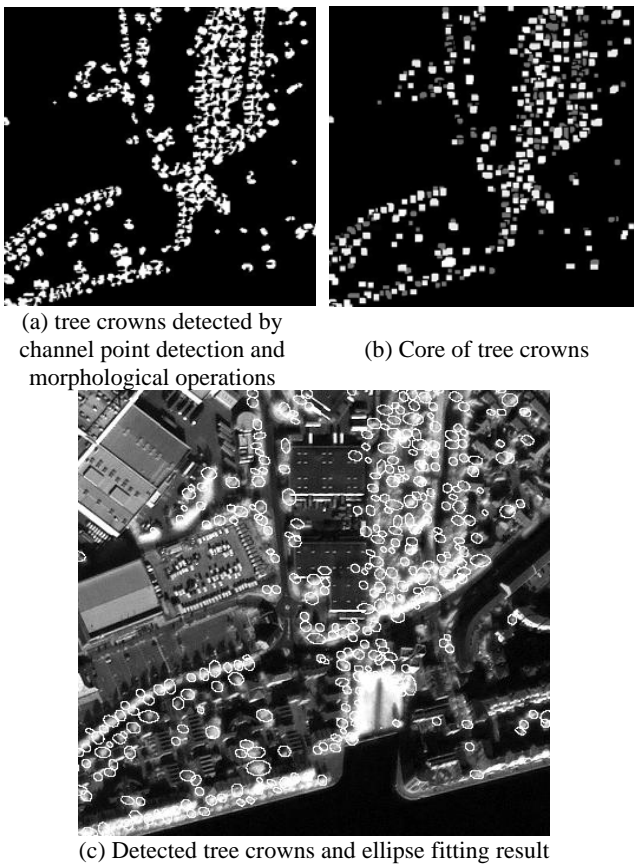

(a) tree crowns detected by channel point detection and morphological operations

(b) Core of tree crowns

(c) Detected tree crowns and ellipse fitting result

Figure 4. Tree detection process by morphological operation and ellipse fitting

4.   RESULTS & EVALUATION

To assess these algorithms, three sub target areas were chosen (Figure 5) which contain different types of building and tree mixtures. Area 1 consists of an area of large buildings on an airbase and has regularly arranged houses. Areas 2 and 3 consist of small housing areas and a few farm buildings so these assess more severe conditions for the building and tree detection.

In all the test areas, there is insufficient ground truth for assessment, so it was necessary to depend on manually extracted landscape objects. Firstly, a building detection ratio was manually assessed. The Building Detection Metrics (Shufelt & McKeown, 1993) scheme is used here to evaluate the accuracy of building outlines compared with GIS data. In the Shufelt and McKeown's scheme, quality assessment factors are defined as below.

$$Detection\_Percentage = 100\frac{TP}{TP+FN}$$

$$Branching\_Factor = \frac{FP}{TP}$$

$$Quality\_Percentage = 100\frac{TP}{TP+FP+FN} \quad (6)$$

where TP is True Positive (Both data sets (detected building and OS data) classify the pixel as being part of a building), TN is True negative (Both data sets classify the pixel as being part of the background), FP is False Positive (Detected data set classifies the pixel as a building, ground truth classifies it as background), FN is False Negatives (Detected data set classifies the pixel as background, ground truth classifies it as a building)
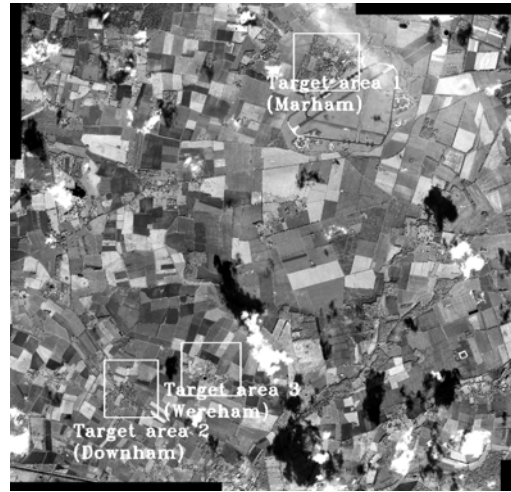

Figure 5. The location of the three test areas Barton Bendish image

In this research, these building detection metrics were not directly available because there were no reliable extracted building polygons available to calculate these metrics. The number of buildings in the limited regions were therefore counted manually and compared with the detected building footprints. The resulting extracted detection ratio (Table 4) is a little bit lower than that of the automatically detected building RoIs based on LiDAR processing which we found in Eastern London where detection percentages are 78-85%. The reason for the lower detection ratio is mainly due to the small building sizes in the target area and errors in the derived DEM, especially in building boundaries. Since the region growing of the DTM construction process climbs across low slopes, building boundaries extracted by unsuitable image matching results, the n-DEM (DEM-DTM) can not be correctly determined. Some part of this error can be solved by supervised classification but the issue cannot yet be fully addressed.

Another error source was the "improper" camera model. Because of the positional errors of the non-rigorous camera model, building heights were determined to be lower than their

actual heights. This phenomenon can be quite significant for such small scale housing areas, and the n-heights lower than the real value must be frequently thresholded in the training vector. Tree detection in very densely wooded areas, which exist over much of the entire IKONOS image, was not attempted because there is no ground truth to assess the results and it is not feasible to assess such dense areas visually. However, the tree detection results around housing areas show a very high accuracy even though trees are frequently "attached" to artificial objects (see Table 5). That's because the detection algorithms exploit the image intensity rather than inaccurate height information.

Normally the quality of stereo 3D products has good quality as observed in Table 6 and Figure 6.

However, the stereo matcher doesn't appear to work well in very homogeneous areas such as ponds, cultivated regions, shadows, roads and cloudy area. Also the highly populated building areas or big forests can result in erroneous DTMs during the gridding process (see "E" area in Figure 6). Therefore the DTMs and n-DEM of these areas are not correctly determined and produce erroneous building detection. Pre-defined masks of urban areas, which may be extracted manually or using pre-existing GIS, and improved shadow and road masks may be necessary to eradicate such error.

Finally, details of tree and building detection results and the DEM in the test areas are shown in Figure 7 so that the quality of products can be visually assessed.

Area	TP	FP	FN	Tree Detection Percentage	Branching Factor	Quality Percentage
1	220	17	74	74.8%	0.077	70.7%
2	146	15	42	77.6%	0.103	71.9%
3	119	26	24	83.2%	0.126	70.4%

Table 4. An assessment of the building detection results

Area	TP	FP	FN	Tree Detection Percentage	Branching Factor	Quality Percentage
1	302	42	32	90.4%	0.139	80.3%
2	220	43	32	83.9%	0.191	74.8%
3	190	32	24	88.7%	0.126	77.2%

Table 5. An assessment of the tree detection results

Area	Max(m)	Min(m)	Mean(m)	Stddev(m)
1	8.533	-3.103	3.793	1.870
2	6.755	0.023	1.546	1.360
3	7.323	-1.011	1.844	1.722

Table 6. Constructed DTM accuracy (Extracted DTM - OS® PANORAMA)

(a) Test area 1

(b) Test area 2, one poor DTM area due to the artefact of stereo height points

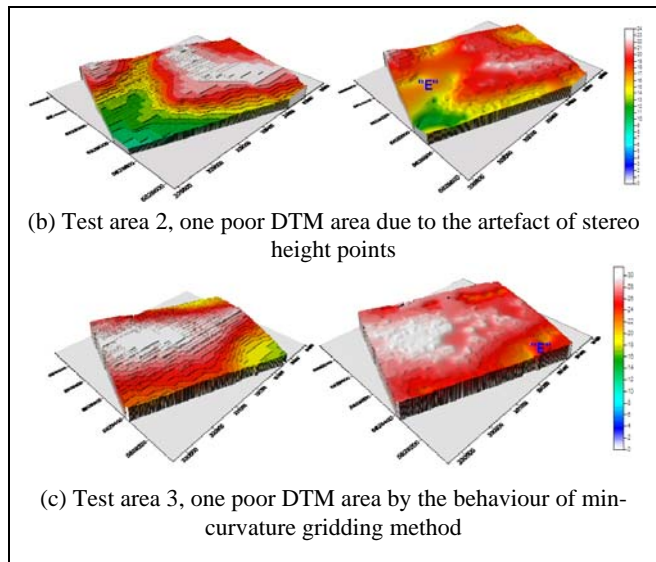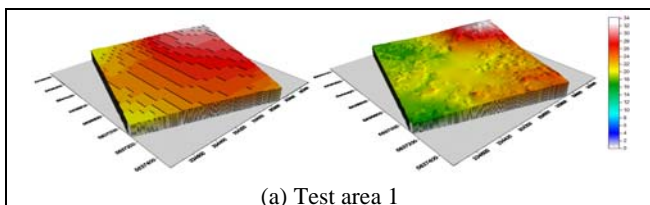(c) Test area 3, one poor DTM area by the behaviour of min-curvature gridding method

Figure 6. The comparison with constructed DTM and OS® PANORAMA DTM ("E" means erroneous DTM)

5. CONCLUSION

Tree and building detection based on stereo IKONOS derived DEMs and multi-spectral classification was performed and validated. The detection ratios of buildings appear a little bit lower than the corresponding ones using IKONOS images and LiDAR DEMs. However, considering the building sizes of Barton-Bendish, which are mostly smaller than the big urban area, the algorithm provides almost the same accuracy using two different 3D range sources, i.e. stereo and LIDAR. A full quantitative quality assessment of the tree crown detection could not be performed due to the absence of ground truth. However, visual inspection indicates partial tree detection algorithm works well. Therefore, the potential of stereo IKONOS for landscape object detection can be concluded to be reasonable.

The most crucial factor for improving the detection ratio is a more efficient stereo matching algorithm to address the building delineation problem and/or an improved generalisation method for building extraction. We are implementing more improved image matcher scheme such as based on dynamic programming. In addition, to compensate for ambiguous object boundary lines, a linear feature matching technique using an eigenvector approach should be employed and then these matched linear features can used for the break-lines in a separate gridding step. In conclusion, the newly developed technique appears to have great potential for automatic landscape database construction solel;y from high resolution stereo satellite image which is much more readily accessible than LiDAR + optical image. We look forward to applying an advanced version of this method to the next generation of high resolution satellites such as Worldview, which will be equipped with finer resolution sensors especially with more spectral bands.

Reference

Adams, R., and Leanne, B. 1994. Seeded Region Growing. *IEEE Transaction Pattern Analysis and Machine Intelligence*. 16(6): 641-647.

Chaud Chaudhuri, B., and Shankar, B. 1989. An Efficient Algorithm for Extrema Detection In Digital Images. *Pattern Recognition Letter*. 10(2): 81-85.

Day, T., Cook, A. C., and Muller, J-P. 1992. Automated Digital Topographic Mapping Techniques for Mars. *International Archives of Photogrammetry and Remote Sensing*. 29(B4): 801-808.

Fraser, C. S., Baltsavias, E., and Gruen, A. 2001a. 3D Building Reconstruction from High-resolution IKONOS Stereo Images. *Procceding of Automatic Extraction of Man-Made Objects from Aerial and Space Images (III), 11-15, June, Ascona, Switzerland*: 13-22.

Fraser, C. S., Hanley, H. B., and Yamakawa, T. 2001b. Sub-metre Geopositioning with IKONOS GEO imagery. *Joint ISPRS Workshop "High Resolution Mapping from Space 2001", 19- 21 September, 2002, Hannover, Germany,* CD-ROM.

Fitzgibbon, A. W., Pilu, M., and Fisher, R. B. 1996. Direct Least Squares Fitting of Ellipses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 21(5): 476-480.

Gruen, A. 1985. Adaptive Least Squares Correlation: A powerful Image Matching Technique. *South Africa Journal of Photogrammetry, Remote Sensing and Cartography*. 14(3): 175-187.

Hanley, H. B., Yamakawa, T., and Fraser, C. S. 2002. Sensor Orientation for High Resolution Satellite Imagery. *Pecora 15/Land Satellite Information IV/ISPRS Commission I/FIEOS 10-15 November 2002, Denver, Colorado US,* http://www.isprs.org/commission1/proceedings/paper/00090.pdf (Accessed 10 January 2005).

Looney, C. G. 2002. Interactive Clustering and Merging with a New Fuzzy Expected Value. *Pattern Recognition*. 35(11): 2413-2423.

Muller, J-P., Kim, J. R., and Kelvin, J. 2004. Automated 3D Mapping of Trees and Buildings and its Application to Risk Assessment of Domestic Subsidence in The London Area. *ISPRS-Congress 2004,12-23 July 2004, Istanbul, Turkey*

Okutomi, M., and Kanade, T. 1992. A Locally Adaptive Window for Signal Matching. *International Journal of Computer Vision*. 7(2): 143-162.

Open GIS Consortium, 2001. Open GIS Abstract Specification - Topic 7: The Earth Imagery Case. https://portal.opengeospatial.org/files/?artifact_id=7467 (Accessed 1 April 2005).

Ordnance Survey, 2003. Land-Form PANORAMA. http://edina.ac.uk/digimap/description/products/panorama.shtml (Accessed 10 January 2005).

Otto, G., and Chau, T. 1989. Region-Growing Algorithm for Matching of Terrain Image. *Image and Vision Computing*. 7(2): 83-94.

Shufelt , J. A., and McKeown, J. D. M. 1993. Fusion of Monocular Cues to Detect Man-made Structures in Aerial Imagery. *Computer Vision, Graphics, and Image Processing*. 57(3): 307-330.

Smith, W. H. F., and Wessel, P. 1990. Gridding with Continuous Curvature Splines in Tension. *Geophysics*. 55, 293-305.
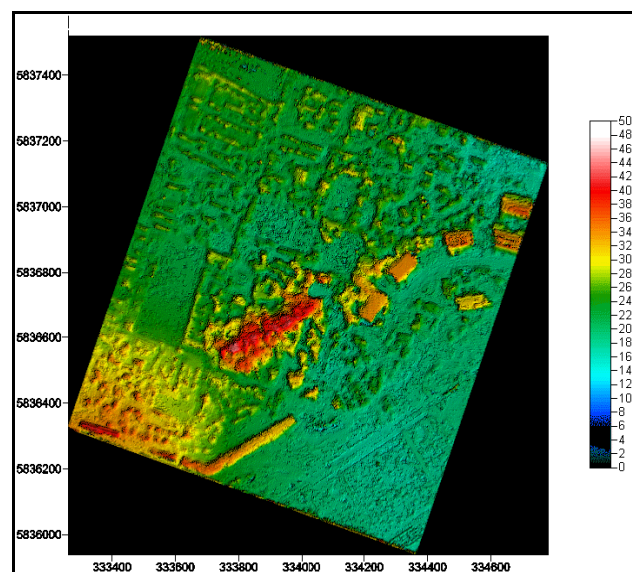
Space Imaging., 2005. http://www.spaceimaging.com/whitepapers_pdfs/IKONOS_Product_Guide.pdf

Wood, J. 1996. The Geomorphological Characterisation of Digital Elevation Models. *Ph.D. Dissertation at the Department of Geography, University of Leicester.*
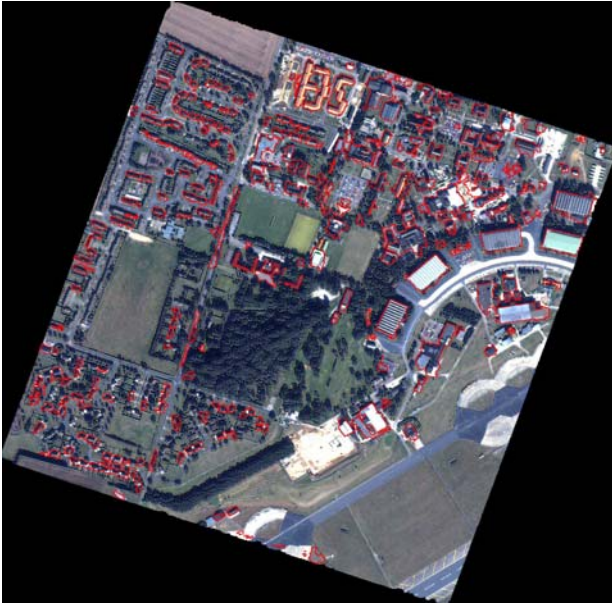
Zitnick, C. I., and Kanade, T. 2000. A Cooperative Algorithm for Stereo Matching and Occlusion Detection. *IEEE Transaction Pattern Analysis and Machine Intelligence 22(7), 675-684.*
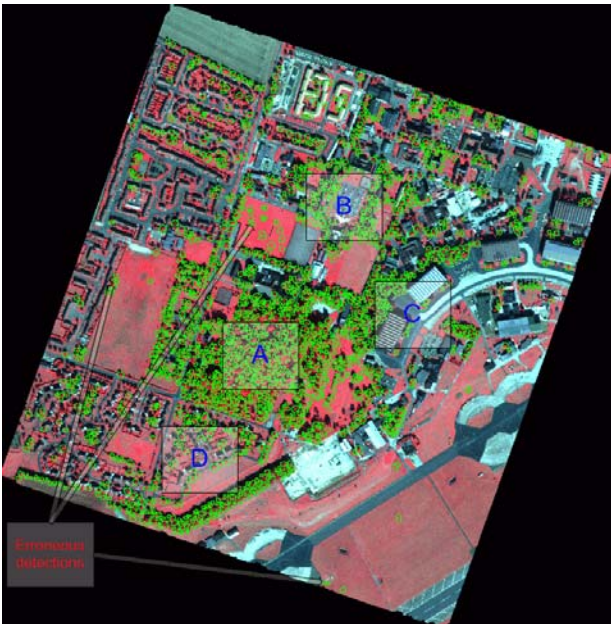
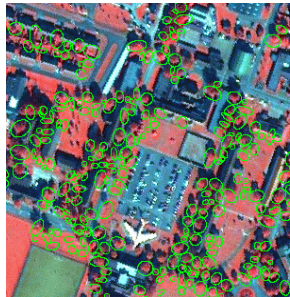## Acknowledgements

(a) Stereo DEM in test area 1

subset C   subset D

(c) Tree detection results and details in test area 1

Figure 7. DEM and object detection results in test area 1

(b) Building detection results in test area 1

subset A   subset B