

3D TERRAIN RECONSTRUCTION BASED ON COMPOUND TECHNIQUES

Yihui Lu¹, Kurt Kubik¹, Mohammed Bennamoun¹ and John Trinder²

¹ Space Centre for Satellite Navigation
School of Electrical and Electronic Systems Engineering
Queensland University of Technology
Brisbane, QLD 4001, Australia

² School of Geomatic Engineering
The University of New South Wales
Sydney NSW 2052
Australia

Keyword: Image matching, 3D reconstruction, Image segmentation, Image interpretation, Object recognition

ABSTRACT

Digital Elevation Models (DEM) produced by digital photogrammetry workstations supply important information on land form for cartographers and Geographic Information Systems (GIS) developers and users. Since the accuracy of maps, orthophotographs and GIS databases must be within a specified range for appropriate analysis of the information and subsequent decision making, an accurate DEM is needed. Current image matching methods for DEM determination only supply a Digital Surface Model (DSM). This means that matching may not occur on the terrain surface, but on the tops of man-made objects such as houses, or on the top of the vegetation. Hence the assumption, that the elevations comprised within a DEM refer to the terrain surface, is incorrect. A 3D terrain reconstruction method using compound techniques is proposed, which will determine terrain elevations more accurately from overlapping digital aerial images and satellite images. It combines image analysis and image matching methods and hence overcomes errors in the DEM caused by matching on tops of trees or man-made objects.

1. INTRODUCTION

A major research area in computer vision and digital photogrammetry is image matching for the reconstruction of a Digital Elevation Model (DEM). This process, which is a fundamental problem in stereo vision, involves the determination of corresponding points in a stereo image pair. From the image coordinates of these corresponding points, their 3D positions can be computed by triangulation, from the known camera geometry, and additional points on the terrain surface can be obtained by interpolation (Zong et al 1992). However, 3D terrain reconstruction from aerial or satellite images will be subject to errors in built-up and treed areas (Tönjes 1996), (Baltsavias et al 1995), (Henricsson et al 1996). In order to obtain a more accurate 3D terrain model, it is necessary to develop better methods to overcome these problems. In this paper, procedures are described that combine image analysis and image matching methods in an attempt to ensure that the elevation points are measured only on the natural terrain surface, and not on the tops of vegetation or man made features such as houses (Miller et al 1994). The paper is organized as follows. Section 2 introduces the proposed system. Section 3 and Section 4 describe the stereo image processing procedure and the single image processing procedure respectively. Section 5 gives experimental results, and conclusions are drawn in Section 6.

2. GENERAL DESCRIPTION OF THE TERRAIN RECONSTRUCTION SYSTEM

Figure 1 illustrates the architecture of the proposed 3D reconstruction system. The goal of this technique is to achieve more accurate reconstruction of elevations from overlapping aerial or satellite images over a wide variety of terrain types and ground cover.

The key functions of data acquisition and pre-processing, are to acquire the images in digital form and improve the output for the subsequent processes by the production of epipolar images from the original left and right images.

The system developed for DEM determination consists of three main parts. Part 1 performs the matching of the stereo image pair, derives a disparity map, and produces a digital surface model (DSM). An analysis of the disparity map then reveals possible house and tree areas. Part 2 applies standard image segmentation and texture analysis techniques to the left image to recognize houses. The method will be extended later to locate

trees. Based on a combination of the 3D information extracted from the disparity map and the 2D image segmentation, the elevations derived in regions which do not appear to represent the terrain surface can be removed from the DSM in Part 3, thus leading to a more accurate DEM. In the case of the houses, the elevations can then be interpolated from the surrounding terrain. Where trees exist, the DSM heights can be reduced by the tree heights.

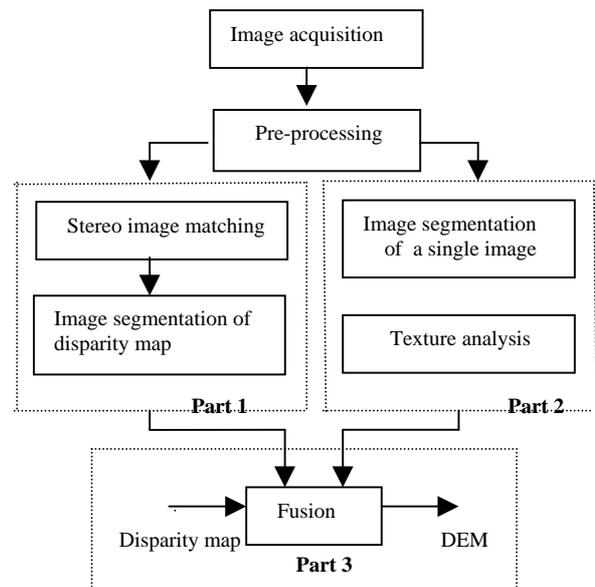


Figure 1 Architecture of the proposed reconstruction system

3. PROCESSING OF STEREO IMAGE PAIR

3.1 Derivation of the Disparity Map

The first step in the recovery of 3D terrain information from overlapping aerial or satellite images is based on the matching of corresponding pixels in the stereo images. From the matched points, the 3D coordinates of a point can be obtained by triangulation using information of the image capturing geometry. Many computational algorithms have been used to solve the

stereo matching problem. Conventional image matching techniques may be classified as either feature-based or area-based. Each of these approaches has advantages and disadvantages. Feature-based matching generally produces good results, is less expensive and is more tolerant of illumination differences and geometric distortions. However, only a few points may be matched in some regions due to the sparsity of the features, which leads to large areas being subjected to inaccurate interpolations (Marr et al 1979), (Barnard et al 1980), (Medioni et al 1985). Area-based matching algorithms can provide denser disparity maps. However, they are intolerant to geometric distortions caused by steep terrain slopes or imaging geometry (Hannah 1989), (Li 1991), (Wu 1995).

In order to produce a dense, reliable matching result, commercial software packages, such as in the Helava and VirtuoZo digital photogrammetry workstations, normally employ hierarchical area-based matching. Since this paper concentrates on the process of recognizing houses and trees in images, and correcting for their effects on derived elevations from image matching, the disparity values obtained from matching in the software package VirtuoZo have been directly used in the subsequent stages of the system in Figure 1. For these developments, a dense sample of points in the disparity map is required in order to avoid some of structures being missed. Hence, a matching grid interval of 5 pixels in column and row directions has been used. The derived disparity map is then interpolated to the same size as the original image for further processing.

3.2 Edge Detection Applied To The Disparity Map

Figure 2 illustrates the stereo image processing procedure of Part 1 in detail.

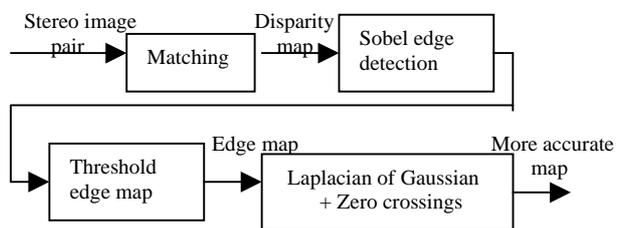


Figure 2 Stereo image processing procedure of Part 1 in Figure 1

Although automatic matching algorithms used by the standard software packages are not able to distinguish between the terrain surface and objects on and above this surface, the output of stereo image matching can supply significant information to identify man-made structures such as houses, and trees. When houses and trees exist in the images, the disparity values of these areas are locally larger, often with discontinuities occurring in the disparity values at the edges of the features (Baltsavias et al 1995), (Henricsson et al 1996). Hence, edge extraction methods, in which relatively distinct changes in grey level properties between two regions in an image are located by changes in local derivatives, are used to define the discontinuities in the disparities values by treating the disparity map as an image. Common methods used to calculate these derivatives are the gradient and Laplacian operators (Gonzales et al 1992). The Sobel gradient operator has the advantage of providing both a differencing and a smoothing effect. Since the derivatives enhance noise, the smoothing effects are a particularly attractive feature of this operator. The Sobel operator is illustrated as follows (KBVision 1996).

1	2	1
0	0	0
-1	-2	-1

(1) Orientation is 0°

2	1	0
1	0	-1
0	-1	-2

(2) Orientation is 45°

The process of edge detection in the disparity map by the Sobel operator involves a repeated convolution of a 3 × 3 operator over the disparity map, by rotating the operator in eight positions about its centre. A maximum response will occur for one of the eight orientations of the operator. This response value and the operator's orientation are stored as values in a 'Sobel image', defining the edges in the disparity map. In order to reduce the effects of noise in extracting the discontinuities, a threshold must be set for recognizing an edge in the disparity map, as a change of disparity. In general, threshold selection is one of the key issues in edge detection (Bennamoun et al 1997). If the threshold level is set too high, it will not permit detection of low magnitude edges, while if it is set too low, it will cause noise to be falsely detected as edges. In this work, the standard deviation (σ) of the variation in the Sobel image magnitudes is used for threshold definition. The pixels in the Sobel image, whose magnitudes satisfy the following threshold formula are deemed to be edge pixels in the disparity image. Equation (1) was obtained by tests to determine an optimal edge image.

$$mag \geq 3 \times \sigma \quad (1)$$

The Sobel operator supplies edges which are more than one pixel wide. To reduce edges to one pixel wide, an edge thinning algorithm is applied to the Sobel image, by computing second derivatives to determine zero-crossings in the image. This is obtained by convolving the disparity image with the Laplacian of a 2D Gaussian function. The pixels which have local maxima in the Sobel image and a zero value for the second derivative can be used to identify the edges in the disparity map and hence the edges of tree and house areas.

The processing of the left image to locate houses, in the next section, involves the steps of: region growing to extract homogeneous regions in the image; an analysis of geometric and radiometric parameters of the regions; determination of the texture of the extracted regions; and the application of morphological operators to confirm the regions.

4. PROCESSING OF SINGLE IMAGE

4.1 Dynamic Region Growing

The left image is processed to separate house and tree areas. Figure 3 illustrates the implementation steps. A dynamic region growing technique is used to determine homogeneous areas in the images in which the intensities of the pixel values are within a given threshold value. The threshold is modified dynamically according to the mean and standard deviation of the pixels in the region while it is being grown, based on Equation (2) (Levine et al 1981). The adaptive threshold will never be larger than the pre-defined threshold (T), but may be smaller.

$$th = (1 - \min(0.8, \text{standard-deviation}/\text{mean})) * T \quad (2)$$

The first region is chosen at the lower left corner of the image and processed until that region can no longer be grown. The next region starts at a pixel that has not been incorporated into the previous region. This process continues until all pixels have been grouped into separate regions which represent homogeneous areas in the input image.

This dynamic region growing technique will locate most homogenous regions, but sometimes the process may cross from one region to the next. To limit this occurrence, the region growing needs to be constrained using an edge detection process. Sobel edge masks are used to produce x and y

direction edge images, which define region boundaries between adjacent pixels in the x and y directions. These x and y edge images are used as input for the region growing and are combined with the dynamic threshold in order to locate the regions more precisely.

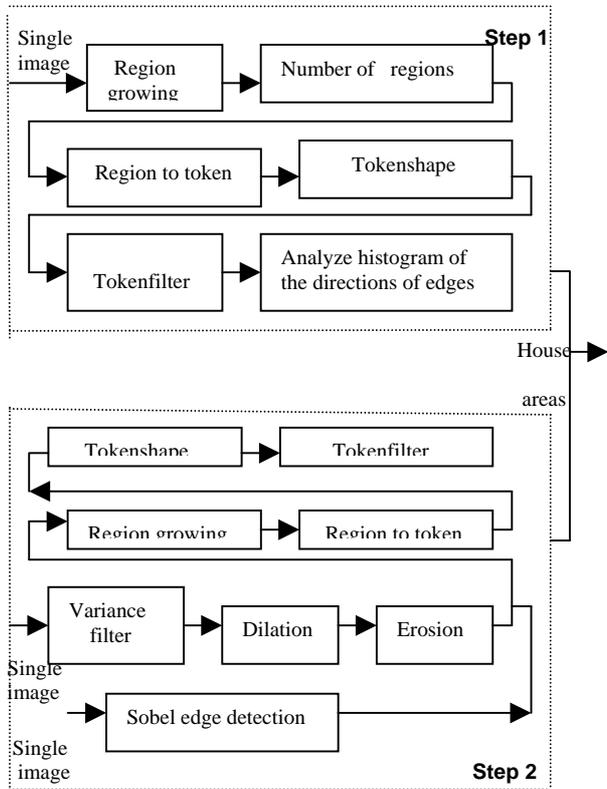


Figure 3 A single image processing procedure of Part 2 in Figure 1

4.2 Analyzing Region Parameters

The number of regions in the image are calculated in the Figure 3. Regions are represented by tokens which also describe the features of that region (KBVision 1996). The task "Region to token" implements the transformation from region to token. "Tokenshape" is used to calculate a series of feature values for the regions, as follows (KBVision 1996):

1)perimeter,

2)intensity_mean, 3)br_to_perimeter= $\left(\frac{perimeter}{2height + 2width}\right)$,

4) $\log_{10}\left(\frac{height}{width}\right)$ and 5)pixel_count. The third

feature value measures region convexity, which is larger for regions with boundaries which double back on themselves. The fourth feature value gives a symmetric measure of aspect ratio.

"Tokenfilter" in Figure 3 is used to filter out extracted regions which are not houses, based on the five feature values for every region. For each image, the minimum sizes of houses, perimeter and pixel_count can be defined. Intensity_mean is based on a special case and helps to extract the houses which have a bright roof. However, houses with dark roofs will have similar grey values as the ground cover, so it is difficult to locate them. Step 2 in Figure 3 can be used to recognize the area of the dark roof house as described below.

4.3 Analyzing the Histogram of Orientations of Edges

After analysing their parameters, most extracted regions can be eliminated, but some regions will be described incorrectly as houses. In order to eliminate these false areas, an analysis is made of the histogram of the orientations of edges of the regions.

Figure 4 illustrates the procedure for distinguishing houses from other objects. There are eight compass directions of region edge pixels in this small test image. Table 1 shows that for regularly shaped houses, the histogram contains a greater number of edge pixels whose orientations are in the horizontal and vertical directions, that is 0° and 90° . This does not occur for all images. However, because the directions of the edge pixels of the houses in this test image are significantly different from those of trees, which are obviously not square, this method can assist in differentiating between houses and trees.

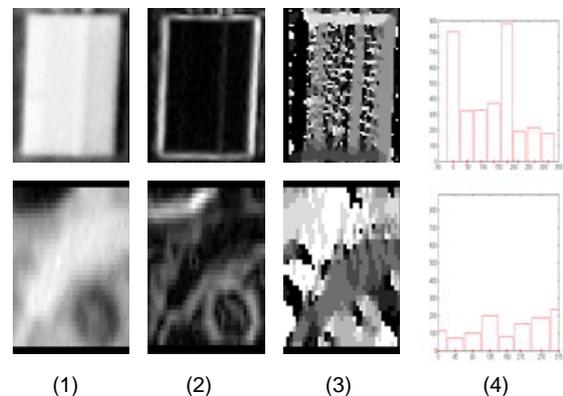


Figure 4 Analysis the difference between two regions (1) original images (2) edge magnitude images (3) edge orientation images (4) histogram of the orientation images

Table 1 The number of different orientation pixels (h: house, t: tree)

	0° *	45°	90° *	135°	180° *
h	832	325	331	371	883
t	115	73	103	201	80
	225°	270° *	315°	Sum of *	Sum of other
h	192	217	179	2263	1067
t	153	187	235	485	662

4.4 Texture Analysis

Locating houses with dark roofs has been addressed in (Vohra et al 1996) by texture analysis. A variance filter is determined from a texture algorithm capable of distinguishing uniform intensity areas in images. Although the dark roofs have similar intensity to the ground cover in the image, they have different textures. The variance filter, which provides a measure of local homogeneity of the intensities in an image, can also be regarded as a non-linear non-directional edge detector (Wilson 1997), (Reed 1993). The variance filter involves replacing a central pixel value with the variance of a specified set of pixel values surrounding it in a window on the image, which does not need to be square. The variance of such a set is given as follows:

$$\bar{x} = \frac{1}{n^2} \sum_{r=1}^n \sum_{c=1}^n x_{rc}$$

$$v_w = \frac{1}{n^2} \sum_{r=1}^n \sum_{c=1}^n (x_{rc} - \bar{x})^2 = \frac{1}{n^2} \sum_{r=1}^n \sum_{c=1}^n (x_{rc}^2) - \bar{x}^2 \quad (3)$$

Where $n \times n$ is the total number of pixels in the window. w is the window in the image, x_{rc} is the value of the pixel at row r and column c in the windows. \bar{x} is the mean of pixel values in the window.

4.5 The Morphological Functions

Morphological transformations are powerful tools for extracting image components, that are useful for representing and describing region shapes (KBVision 1996), (Kanai 1998). Dilation combines two image sets using vector addition of set elements, while erosion combines two sets by vector subtraction. Examples of morphological operators are given in Figure 5.

After calculating and thresholding the variance for each pixel derived by equation (3), the area of the dark roof can normally be recognized. If some of the points in the dark roofs have a higher variance, some small sections may not be correctly assigned as roof pixels. A dilation operation can be used to fill in these small areas. Since some of the boundary pixels of dark roof areas may be connected with other regions, an erosion operation can be used to separate them from other regions. Dark roofs can thus normally be extracted.

By combining the results of morphological transformation with the edges derived from the Sobel edge detection, a more accurate region boundary can be obtained. As shown in step 2 in Figure 3, the subsequent operations are similar to the ones described in step 1. After steps 1 and 2, the final house regions can be obtained. Figure 6 illustrates the results of processing step 2 in Figure 3.

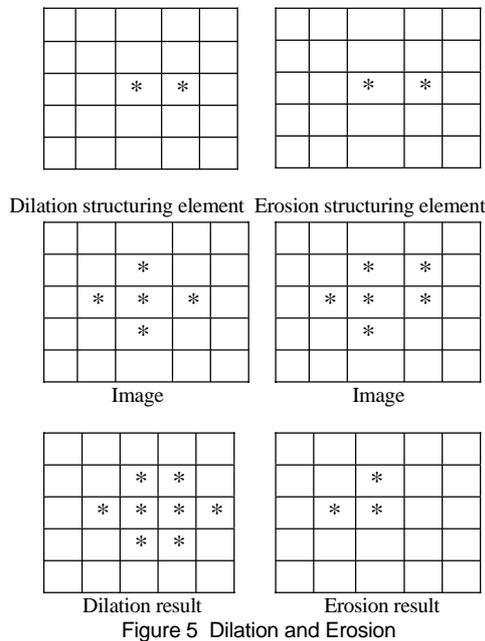
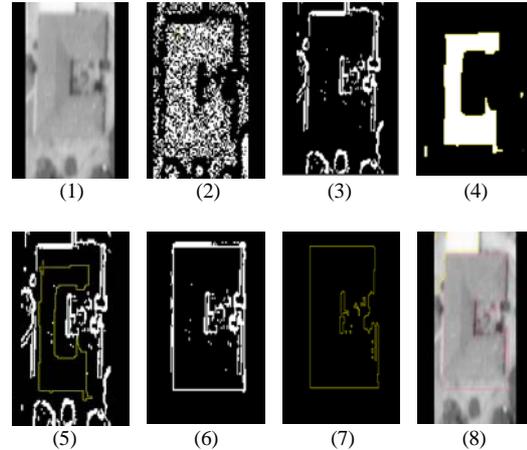


Figure 5 Dilation and Erosion



(1)Image of dark roof house (2) Result of threshold of variance (3) Edges detected by the Sobel operator (4) Result of the morphological transformation (5) Combined result of (3) and (4) (6) New boundary location for dark roof (7) Region of dark roof (8) Region overlaid on the image

Figure 6 Results by step 2 in Figure 3

5. TEST AND RESULTS

Figure 7 illustrates a pair of aerial images with 630×714 pixels in the row and column directions respectively. The scale of image is 1:3437. The flying height is 519 metre. The focal length of the camera is 153mm and pixel size is $100 \mu m$. Figure 8 illustrates the disparity map obtained from stereo image matching in the digital photogrammetric workstation. The disparity map is further processed to obtain the house and tree areas illustrated in Figure 9, using the method described in Section 3.

The left image has then been processed to locate houses and to separate them from trees by the method described in Section 4. Figure 10 shows the regions obtained by region growing in step 1 in Figure 3. Based on the five feature values of regions, most of the houses can be defined, as illustrated in Figure 11. For each region in Figure 11, the corresponding small image region in the original image can be identified.

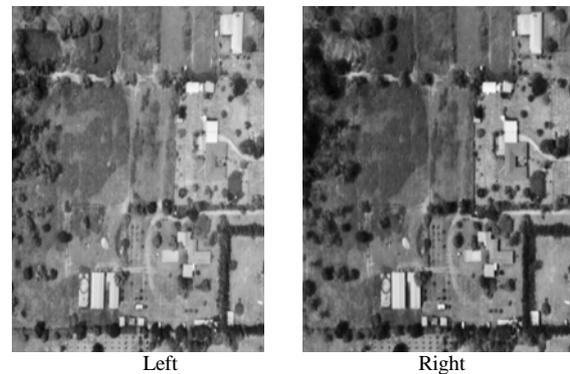


Figure 7 Stereo image pair



Figure 8 Disparity map

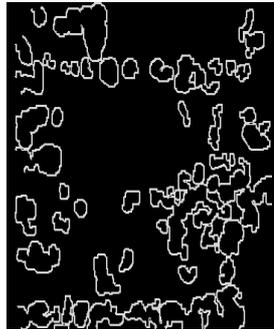


Figure 9 outlines of house and tree areas

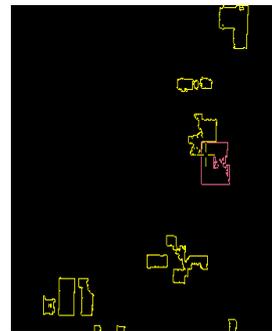


Figure 14 Final delineation of houses

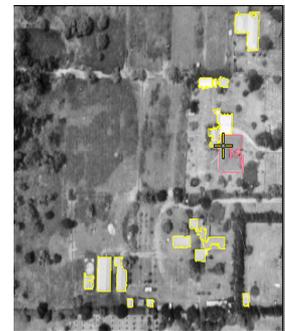


Figure 15 Extracted roofs overlaid on the image

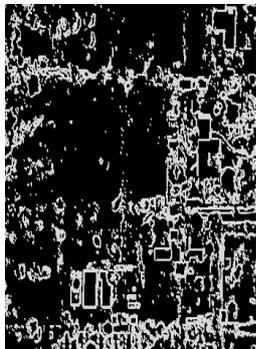


Figure 10 Results of region growing

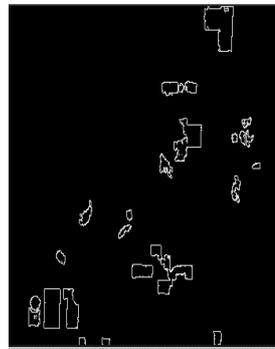
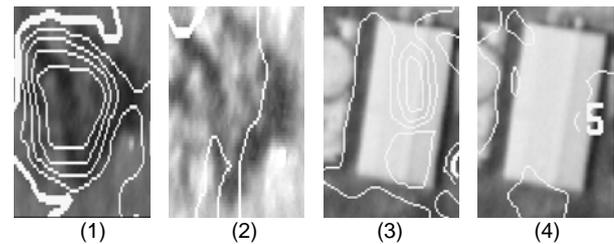


Figure 11 Definition of regions after the application of region parameters

Using the compound information from the analysis of image matching and 2D image segmentation, some digital elevation points which were initially located on the tops of houses and trees have been interpolated onto the ground. Figure 16 illustrates the contour changes in these areas. The DEM was checked for elevation accuracy by comparing the DEMs derived from the digital photogrammetry workstation Virtuoso, and



(1) Contour in tree area from matching (2) Corrected contours (3) Original contours in house area (4) Corrected contours

Figure 16 Contours obtained from the image matching and corrected contours based on the method described in the paper

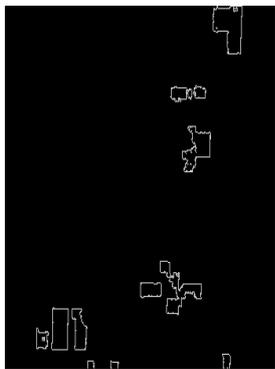


Figure 12 Houses obtained by step 1 in Figure 3

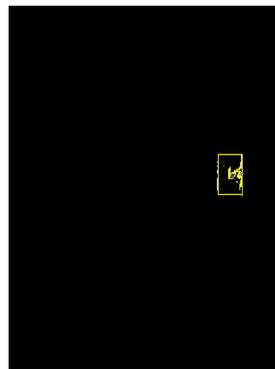


Figure 13 The dark roof area

that derived by the method described in this paper, with a DEM obtained by an human operator on a Zeiss Planicomp analytical plotter. The comparisons of results are shown in Table 2. DZ are the differences between DEMs. Since the operator of the Planicomp analytical plotter could not supply enough comparison points in house areas, DEMs in these areas can not be compared with the true ground values. However, Table 2 demonstrates that the combined approach described in this paper produces better results than the image matching only approach.

Table 2 The comparison of DEMs between the plotter and two other methods

Type	Our proposed method	Virtuozo
No. of points in DEM	2581	2580
Max(DZ) (metre)	5.900	16.40
σ (DZ) (metre)	1.075	1.352

Based on an analysis of the histogram of the orientations of edges, areas whose edges are not considered to be those of houses can be eliminated. The extracted house areas are again illustrated in Figure 12. Figure 13 shows the area of dark roof extracted by step 2 in Figure 3, while Figure 14 shows the final determination of the houses. Figure 15 is the image in which the final extracted house areas are overlaid on the original image. This figure shows that the houses are well delineated.

Figures 17 and 19 illustrate the DEM and 3D perspective view derived directly from stereo image matching produced by Virtuoso. A more accurate DEM and 3D perspective view produced by the method in this paper are shown in Figures 18 and 20.

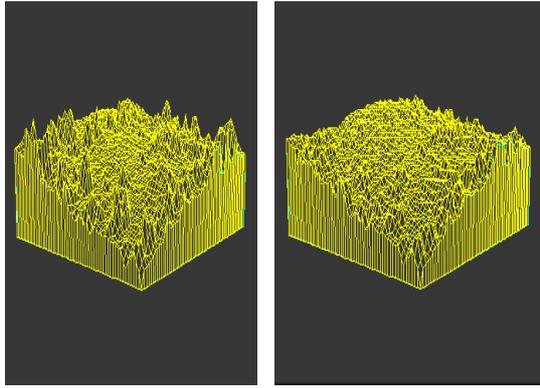


Figure 17 DEM from matching Figure 18 DEM from the combined method

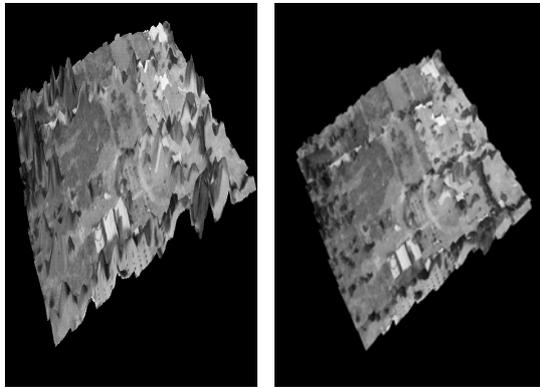


Figure 19 3D perspective view from matching Figure 20 3D perspective view derived by the combined method

6. CONCLUSION

The method described in this paper combines image matching and image analysis methods, which enables the location of most of the house and tree areas in the test images. These areas are important information for 3D terrain reconstruction and ensure that points are only measured on the natural terrain. The method leads to more accurate determination of elevations from overlapping digital aerial images than the DSM determined only by image matching, since it avoids errors caused by man-made or natural surface features.

The method can also locate dark roofs. One disadvantage is its inability to exactly locate the boundary of dark roofs in cases when the roof of a house is not square, or when houses and trees are close to each other, since the detected boundaries of the house and tree areas overlap. Further research is needed to find a more reliable method for these areas. The method will also be tested on other scales and different images.

References

Baltsavias E., Mason S. & Stallmann D. (1995). Use of DTMs/DSMs and Orthoimages to Support Building Extraction. Automatic Extraction of Man-made Objects from Aerial and Space Images. Birkhauser Verlag, Basel, pp 199-210.

Barnard S. T. & Thompson W. B. (1980). Disparity Analysis of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-2(4):333-340.

Bennamoun M. & Boashash B. (1997). A Probabilistic Approach for Automatic Parameters Selection for the Hybrid Edge Detector. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E80-A(8): 1423-1429.

Gonzalez R. C. & Woods R. E. (1992). *Digital Image Processing*. Addison-Wesley, U. S. A.

Hannah M. J. (1989). A System For Digital Stereo Image Matching. *Photogrammetric Engineering and Remote sensing*, 55(12): 1765-1770.

Henricsson O., Bignone F., Willuhn W., Ade F., & Kuebler O. (1996). Project AMOBE: Strategies, Current Status and Future Work. *International Archives of Photogrammetry and Remote Sensing*, 31(3):321-330.

Kanai Y. (1998). Image Segmentation Using Intensity and Color Information. *Visual Communications and Image Processing'98*, 3309:709-720.

KBVision System Task Reference Manual (1996). Amerinex Applied Imaging, Inc.

Levine M. & Shaheen S. (1981). A Modular Computer Vision System for Image Segmentations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI 3(5): 540-554.

Li M. (1991). Hierarchical Multipoint Matching. *Photogrammetric Engineering and Remote Sensing*, 57(8): 1039-1047.

Marr D. & Poggio T. (1979). A Computational Theory of Human Stereo Vision. *Proceedings of Royal Society of London*, B204, pp. 301-328.

Medioni G. & Nevatia R. (1985). Segment-based stereo Matching. *Computer Vision, Graphics, and Image Processing*, 31: 2-18.

Miller S., Walker S. & Walsh M.C. (1994). Digital Photogrammetry at the Ordnance Survey of Ireland. Report by Helava and Associates.

Reed T. R. (1993). A review of Recent Texture Segmentation and Feature Extraction Techniques. *Computer Vision, Graphics and Image Processing: Image Understanding*, 57(3):359-372.

Tönjes R. (1996). Knowledge Based Modelling of Landscapes. *International Archives of Photogrammetry and Remote Sensing*, 31(3):868-873.

Vohra V. K. & Dowman I. J. (1996). Automatic Extraction of Large Buildings From High Resolution Satellite Images For Registration with a Map. *International Archives of Photogrammetry and Remote Sensing*, 31(3):903-905.

Wilson P. A. (1997). Rule-Based Classification of Water in Landsat MSS Images Using the Variance Filter. *Photogrammetric Engineering & Remote Sensing*, 63(5):485-491.

Wu X. (1995). Grey-based Relaxation for Image Matching. *3rd Conference on Digital Image Computing: Techniques and Applications*, December, 1995, Brisbane, Australia.

Zong J., Li J. & Schenk T. (1992). Aerial Image Matching Based on Zero-Crossings. *International Archives of Photogrammetry and Remote Sensing*, 29(3):144-150.