STEREO IMAGE MATCHING USING ROBUST ESTIMATION AND IMAGE ANALYSIS TECHNIQUES FOR DEM GENERATION

Yihui Lu [1,2] and Kurt Kubik [1]

[1] Department of Geographical Sciences and Planning
University of Queensland, QLD 4072 Australia

[2] Department of Computer Science and Electrical Engineering
University of Queensland, QLD 4072 Australia

KEY WORDS: Image matching, computer vision, image understanding, DEM

ABSTRACT

Digital Elevation Models (DEM) produced by digital photogrammetry workstations are often used as a component in complex Geographic Information Systems (GIS) modeling. Since the accuracy of GIS databases must be within a specified range for appropriate analysis of the information and subsequent decision making, an accurate DEM is needed. Conventional image matching techniques may be classified as either area-based or feature-based methods. These image matching techniques could not overcome the disparity discontinuities problem and only supply a Digital Surface Model (DSM). This means that matching may not occur on the terrain surface, but on the top of man-made objects such as houses, or on the top of the vegetation. In order to get more accurate DEM from overlapping digital aerial images and satellite images, a 3D terrain reconstruction method using compound techniques is proposed. The area-based image matching method is used to supply dense disparities. Image edge detection and texture analysis techniques are used to find houses and tree areas. Both these parts are robustified in order to avoid outlyers. The final DEM comes from the two parts of image matching and image analysis 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. However, 3D terrain reconstruction from aerial or satellite images will be subject to errors in built-up and treed areas [Baltsavias et al 1995,Henricsson et al 1997 & Tonjes 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 top of vegetation or man made features such as houses. 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

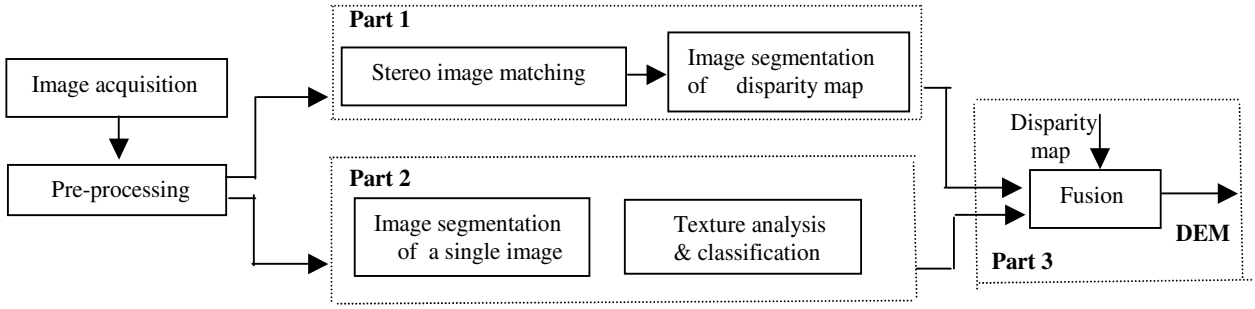


Figure 1 Architecture of the proposed reconstruction system

left image to recognize houses and 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.

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 scarcity of the features, which leads to large areas being subjected to inaccurate interpolations. Area-based matching algorithms can provide denser disparity maps. However, they are intolerant to geometric distortions caused by steep terrain slopes or imaging geometry.

In order to produce a dense, reliable matching result, the hierarchical area-based stereo image matching using robust estimation was been employed. 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 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.

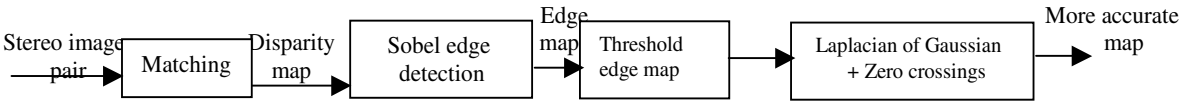


Figure 2 Stereo image processing procedure of Part 1 in Figure 1

Although automatic area-based matching algorithms 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. 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 (Gonzalez & Woods 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.

4 PROCESSING OF SIGLE IMAGE

4.1 Single Image Processing for House Extraction

The left image is processed to separate house and tree areas. Figure 3 illustrates the implementation steps.

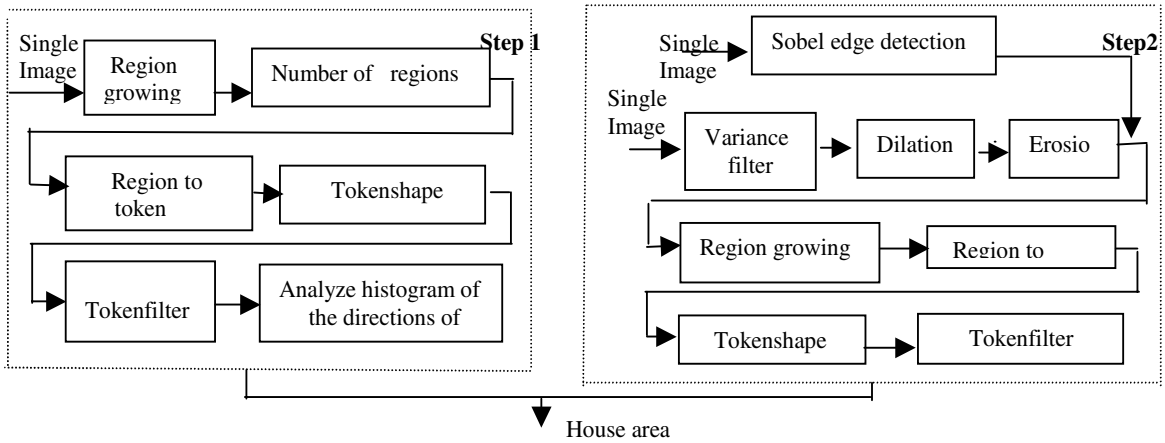


Figure 3 A single image processing procedure of Part 2 in Figure 1

4.1.1 Dynamic region growing

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 (KBVision 1996). The adaptive threshold will never be larger than the pre-defined threshold (T), but may be smaller. It is th = (1-min(0.8, standard-deviation/mean))*T. 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.

4.1.2 Analysing 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:

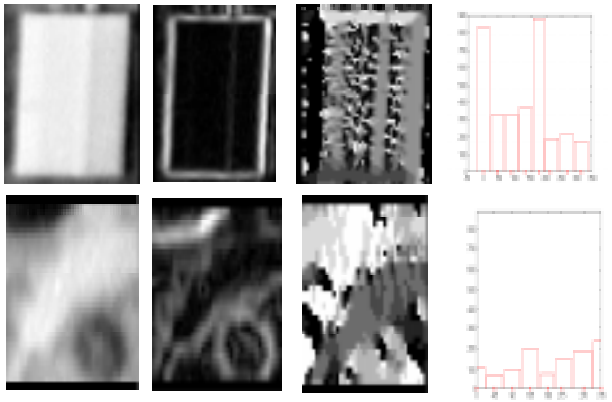
1)perimeter, 2)intensity_mean, 3)br_to_perimeter= $(\frac{perimeter}{2height + 2width})$ 4)log_h_to_w= $\log_{10}(\frac{height}{width})$ 5)pixel_count.

"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.

4.1.3 Analysing 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 edges whose orientations are mutually orthogonal such as here at 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.



(1) (2) (3) (4)

(1)original images (2) edge magnitude images (3) edge orientation images (4) histogram of the orientation images

Figure 4 Analysis the difference between two regions

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

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

4.1.4 Texture analysis

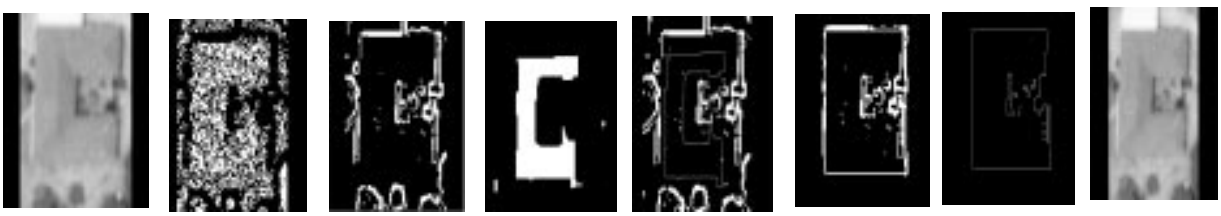
A variance filter is determined from a texture algorithm capable of distinguishing uniform intensity areas in images. Although the dark roofs have similar intensity to the ground cover in the image, they have different textures. The variance filter, which provides a measure of local homogeneity of the intensities in an image, can also be regarded as a non-linear non-directional edge detector (Wilson 1997). 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:

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

Where n *x* 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.

4.1.5 The morphological functions

Morphological transformations are powerful tools for extracting image components, that are useful for representing and describing region shapes. Dilation combines two image sets using vector addition of set elements, while erosion combines two sets by vector subtraction.



(1) (2) (3) (4) (5) (6) (7) (8)

(1)Image of dark roof house (2) Result of thresheld 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 5 Results by step 2 in Figure 3

After calculating and thresholding the variance for each pixel derived, 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 5 illustrates the results of processing step 2 in Figure 3.

4.2 Single Image Processing for Delineation Tree

4.2.1. Image segmentation

The image is processed to recognize tree areas. Figure 6 illustrates the implementation steps. Using combined results from step1 and step2, delineation of tree areas can be more accurate than either step1 or step 2.

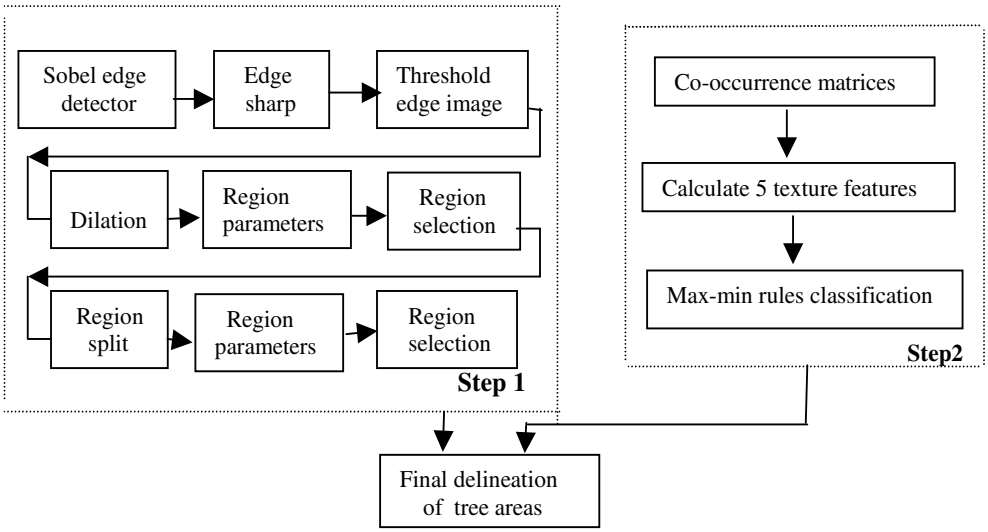


Figure 6 Image segmentation and classification for delineation tree areas

The steps of image segmentation are illustrated in the step1 in Figure 6. The process of edge detection, morphological functions and analyzing region parameters are described as in 4.1. Region split is an important step in delineation tree areas, since trees are usually close to each other or to other objects, some regions remain connected. The “region split” function is used to split the connected regions and eliminate undesirable regions. Detailed can be found in (KBVision 1996).

This task splits the regions into new smaller homogenous regions based on peak-valley analysis of the associated pixel intensity histogram. For each region, a pixel intensity histogram is created. Only pixels that lie within the region are used to compute the histogram. The task automatically selects values and use them to threshold the pixels in the region. Contiguous pixels that fall within the same threshold boundaries form new regions

4.2.2 Texture analysis and image classification using co-occurrence matrices

Co-occurrence matrices has been described in image processing literature by a number of names including gray-tone spatial-dependence matrices (Haralick et al 1973 & Conners & Harlow 1980). It has been widely used in texture analysis and classification. The co-occurrence matrix in it core is a two-dimensional histogram of the occurrence of pairs of

intensity values in a given spatial relationship. Co-occurrence matrices are based on the relative frequencies $p(i, j)$ with which two pixels with a specified separation occur in the image, one with gray level I and another with gray level J. The separation is usually specified by distance vector $\phi(d,\alpha)$. Pixel distance d and angular orientation α are parameters of a particular co-occurrence matrices. Using different parameters, several matrices can be derived. The matrices obtained are then used for extraction of texture features.

For the image of N gray level, co-occurrence matrices cm can be obtained by estimating the pairwise statistics of pixel intensity. The size of cm is determined by the number of gray level N in the input image. The matrices cm are a functions of the angular relationship between the pixels as well as a function of the distance between them. The matrices can be illustrated as following: $cm(d,\alpha) = [p(i, j \mid d, a)]$. After introducing the symmetry (Burns & Smith 1996), we can only consider the angular α up to 180° rotation. The value d normally be chosen as 1. An example of a 4x4 image with four gray levels and the computation of the co-occurrence matrices for $d=1$ and α varying from 0° to 135° by 45° increments are shown in Figure 7:

0	0	1	1
0	0	1	1
0	2	2	2
2	2	2	3

(a) image of 4 gray levels

(Gray levels)	0	1	2	3
0	#(0,0)	#(0,1)	#(0,2)	#(0,3)
1	#(1,0)	#(1,1)	#(1,2)	#(1,3)
2	#(2,0)	#(2,1)	#(2,2)	#(2,3)
3	#(3,0)	#(3,1)	#(3,2)	#(3,3)

(b) general form of any co-occurrence matrix (#(i,j) is the number of times gray level i and j have been neighbors)

$$cm(1,0^\circ) = \begin{bmatrix} 4 & 2 & 1 & 0 \\ 2 & 4 & 0 & 0 \\ 1 & 0 & 6 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix} \quad cm(1,45^\circ) = \begin{bmatrix} 4 & 1 & 0 & 0 \\ 1 & 2 & 2 & 0 \\ 0 & 2 & 4 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad cm(1,90^\circ) = \begin{bmatrix} 6 & 0 & 2 & 0 \\ 0 & 4 & 2 & 0 \\ 2 & 2 & 2 & 2 \\ 0 & 0 & 2 & 0 \end{bmatrix} \quad cm(1,135^\circ) = \begin{bmatrix} 2 & 1 & 3 & 0 \\ 1 & 2 & 1 & 0 \\ 3 & 1 & 0 & 2 \\ 0 & 0 & 2 & 0 \end{bmatrix}$$

Figure 7 Co-occurrence matrices for four given distance vectors (taken from Haralick[1973])

In texture classification, individual elements of the co-occurence are rarely used. Instead, features are derived from the matrix. A large number of textural features have been proposed starting with the original fourteen features described by Haralick, however only some of these are in wide use. The features we used are listed as following:

1. Inverse Difference Moment $f_1 = \sum_{i,j} P(i, j)/1+(i-j)^2$

2. Contrast $f_2 = \sum_{i,j} \delta_{ij}^2 P(i, j)$

3.Entropy $f_3 = -\sum_{i,j} P(i, j) \log P(i, j)$

4. Correlation $f_4 = \sum_{i,j} (i-\mu_i)(j-\mu_j)P(i, j)/\sigma_i\sigma_j$

5. Energy (angular second moment) $f_5 = \sum_{i,j} P(i, j)^2$

μ_i and μ_j are the means and σ_i and σ_j are the standard deviations of i and j respectively. Rotating α, there are 4 values for every texture feature. Then the minimum and maximum values of the texture features can be obtained. Since we only want to find tree areas, the number of samples for tree category is small. We use a min-max decision rule for classification of the image based on their texture features. The procedure is repeated for all the image blocks in the image. The decision rule is described by the following equation. If the equation is satisfied, the processed image block j can be assigned as catalogue k.

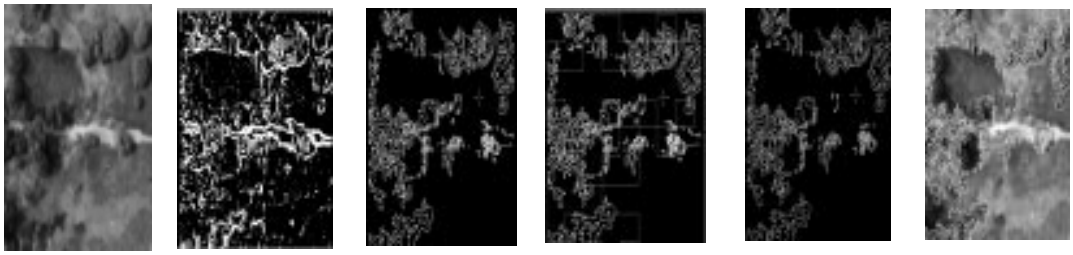
$$\prod_{n=1}^{N} \left(\frac{1}{a_{nk} - b_{nk}}\right) \geq \prod_{n=1}^{N} \left(\frac{1}{a_{nj} - b_{nj}}\right) \qquad (2)$$

b_{nk}, a_{nk}, b_{nj} and a_{nj} define the minimum and maximum texture feature values of the training samples and processed image blocks. n is the number of texture features

Figure 8 illustrates the delineation tress areas using the steps in Figure 6. Using image segmentation method described in Section 4.2.1, most of the tree areas can be delineated. Some areas obtained are not correct. Using the texture analysis and

image classification method introduced in Section 4.2.2, the original image can be classified to supply image blocks which are possible tree areas.

The results from two steps are overlaid together and shown in Figure 8(4). Combing the results from two steps, the final tree areas can be illustrated in Figure 8(5). Some small areas which are extracted incorrectly, are eliminated. Figure 8(6) is the extracted tree areas overlaid on the image.



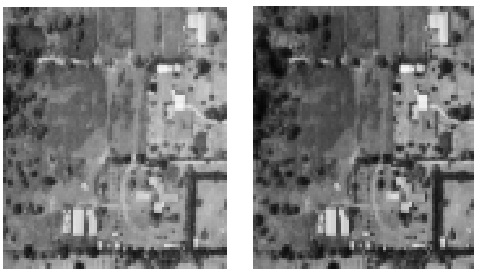
(1) (2) (3) (4) (5) (6)

(1) Testing image area (2) threshold edge image(3) Tree areas from segmentation (4) Two results from Figure 6 (5)Final delineation of tree areas (6) Extracted tree areas overlaid on image

Figure 8 Delineation tree areas using steps in Fig.7

5 TEST AND RESULTS

Figure 9 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 μm . Figure 10 illustrates the disparity map obtained from stereo image matching using robust estimation. The disparity map is further processed to obtain the house and tree areas illustrated in Figure 11, using the method described in Section 3.



Left Right

Figure 9 Stereo image pair

Figure 10 Disparity map

Figure 11 Outlines of house and tree areas

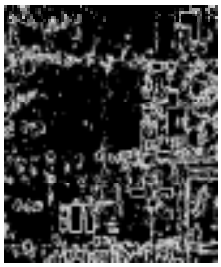


Figure 12 Results of region growing

Figure 13 Definition of regions after the application of region parameters

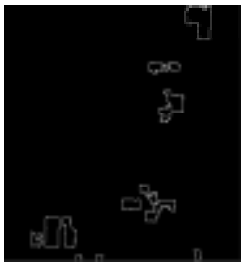


Figure 14 Houses obtained by step 1 in Figure 3

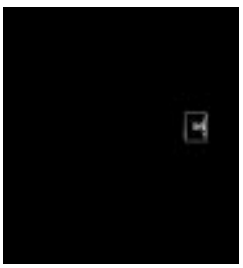


Figure 15 The dark roof area

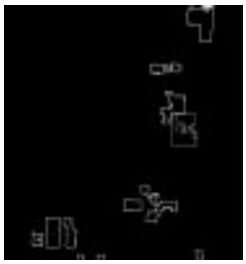


Figure 16 Final delineation of houses

Figure 17 Extracted roofs overlaid on the image

The left image has then been processed to locate houses and to separate them from trees by the method described in Section 4.1. Figure 12 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 13. For each region in Figure 13, the corresponding small image region in the original image can be identified.

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 14. Figure 15 shows the area of dark roof extracted by step 2 in Figure 3, while Figure 16 shows the final determination of the houses. Figure 17 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.

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. Figures 18 and 20 illustrate the DEM and 3D perspective view derived directly from normal stereo image matching. A more accurate DEM and 3D perspective view produced by the method in this paper are shown in Figures 19 and 21.

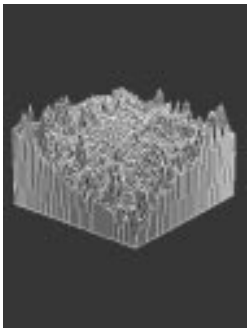


Figure 18 DEM from matching

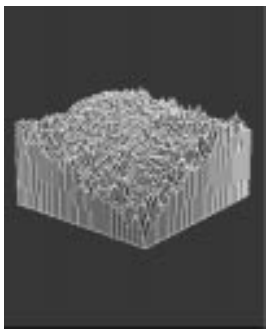


Figure 19 DEM from the combined method

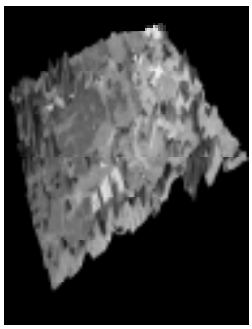


Figure 20 3D perspective view from matching

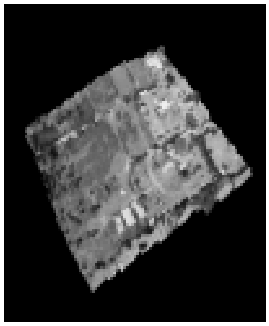


Figure 21 3D perspective view derived by the combined method

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. The image segmentation and classification methods overcome the weakness of co-occurrence matrices that is it does not consider the shapes of gray level primitives. These extracted house and tree 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. The disadvantage is its inability to exactly locate the boundary of dark roofs in cases when the roof of a house is not of regular shape. Since the classification result of co-occurrence matrices are dependent on chosen training sample and the size of the processed image block, further research is needed to find a more reliable method for image classification. 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.

Burns, I. and Smith, G. (1996) MeasTex Version 1.0: A Framework for Measuring the Perfomance of Texture Classification Algorithms. University of Queensland, http://www/cssip.elec.uq.edu.au/~guy/meastex/meastex.html.

Conners, R. W., Harlow, C. A. (1980). A Theoretical Comparison of Texture Algorithms*, IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-2, No. 3, pp.204-222.

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

Haralick, R. M., Shanmugam, K. S., and Dinstein, I. (1973) Textural Features for Image Classification. *IEEE International Conference on Systems, Man, and Cybernetics*, Vol. SMC-3, No.6, pp610-621.

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.

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

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

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.