MVA'94 IAPR Workshop on Machine Vision Applications Dec. 13-15, 1994, Kawasaki

AUTOMATED BUILDING HEIGHT EXTRACTION AND BUILDING DETECTION FROM HIGH RESOLUTION AERIAL AND SPACE IMAGERY

Taejung Kim Jan-Peter Muller
Department of Photogrammetry and Surveying
University College London
Gower St., London WC1E 6BT, U.K.
INTERNET address: kim@ps.ucl.ac.uk, jpmuller@ps.ucl.ac.uk

ABSTRACT

An automatic height extraction algorithm using a stereoscopic approach and a building detection algorithm using a monoscopic approach are introduced. The automated height extraction is done using a pyramidal matching strategy and a silhouette mask made from linear elements of buildings. A pyramidal matching algorithm with a tile-based seed point selection strategy between levels and automatic seed points at the coarsest level matching was developed. A silhouette image of buildings is used to remove possible blunders around building boundaries. The accuracy of the algorithm is assessed quantitatively and is shown to have $4m$ RMS. errors for the area studied.

A novel building detection algorithm is proposed based on a graph constructed from lines and line relations. Building hypotheses are generated by finding closed loops in the graph by a depth-first graph traversal algorithm. Building hypotheses are then verified after merging and removing false building hypotheses by using shadow lines and vertical lines. This algorithm was tested with $15cm$ resolution air-borne image and $2m$ resolution space-borne image and shown to work successfully.

INTRODUCTION

In the late 1980s an automated system for 3D co-ordinate extraction, the UCL 3D Image Maker [1] was developed for close-range [2] and spaceborne stereo [1]. Several attempts were made at UCL [3] to enhance the 3DIM system for high resolution aerial photography including the use of pyramidal matching and randomised seed points. This paper describes work on the use of "intelligent" segmentation techniques to isolate regions of potential blunders in stereo imagery as well as a quantitative analysis of the accuracy of height measurement using automated stereo vision.

Automated segmentation of buildings from aerial and high resolution ($2m$) Russian space imagery (DD5) using a line relation graph and a depth-first graph traversal algorithm will be described as well as a quantitative analysis of the performance of this new algorithm for an urban and urban-rural boundary.

AUTOMATIC HEIGHT EXTRACTION

High resolution urban area imagery have many distinctive characteristics compared with low resolution imagery that make many approaches for low resolution imagery inappropriate. Buildings create abrupt height discontinuities, occlusions, and shadows. Moving vehicles are also detectable in urban area imagery. Previous experiments with the region growing Gruen's adaptive least squares correlation(ALSC) stereo matching algorithm("Gotcha") [4, 5] shows that buildings make isolated regions for region growing [6]. As the "Gotcha" needs at least one initial seed point in an isolated region, a large number of seed points are required for urban area stereo matching.

A pyramidal matching algorithm was implemented as a possible solution. Pyramidal matching solves the problem by utilising the results of lower resolution matching as initial seed points in higher resolution matching. In this way, pyramidal matching facilitates well distributed initial seed points which cover as many isolated regions as possible. Pyramidal matching also reduces the magnitude of height discontinuities and the number of isolated regions by averaging down the resolution of the image. At the lowest (coarsest) level, matching can be performed successfully and in the higher level, matching can be enhanced by utilising the results of lower level matching.

Another advantage of pyramidal matching is the use of automatic seed points, which enables the development of an automated system. Automatic seed points are generated by assuming a constant disparity value at the coarsest level and verified by non-region growing ALSC matching. As disparity range at the coarsest level will be small, actual conjugate points can be found after pre-matching. They are used as seed points to start pyramidal matching.

An appropriate control strategy between levels should be used to reduce the problem of blunder propagation. [6]

In our pyramidal matching, a tile-based selection strategy was adopted as it can achieve well distributed seed points by defining tiles spread over the image space and selecting one match point with best accuracy within a tile as a seed point for the next level. The maximum eigenvalue of the covariance matrix of estimation is used as a measure of accuracy, which should be zero for an ideal matching.

Strong edge elements presented in the imagery were used to produce the silhouette image of buildings. The Canny-Petrou-Kittler(CPK) filter [7, 8] was used to extract edge elements. A Hough transformation based on Duda and Hart [9] and a connected edge labelling algorithm [6] were applied to the edge elements and their results were combined to select linear elements from building boundaries. A DEM of the linear elements was created by assigning a 3D coordinate to a point in a linear element based on the results from pyramidal matching. A silhouette image of buildings was then created by thresholding height to remove the ground level elements and applying a hole-filling operation to recover building structures from the wire-frame-like building boundaries.

A Digital Elevation Model (DEM) from pyramidal matching was found to have high magnitude of errors near building boundaries. A silhouette image of buildings were used to overcome these errors. This is because it can locate buildings more precisely as it has been generated from linear elements. A silhouette image is used as a mask to a DEM from pyramidal matching to filter out the erroneous building level points. Fig 1 summarises the overall procedure.

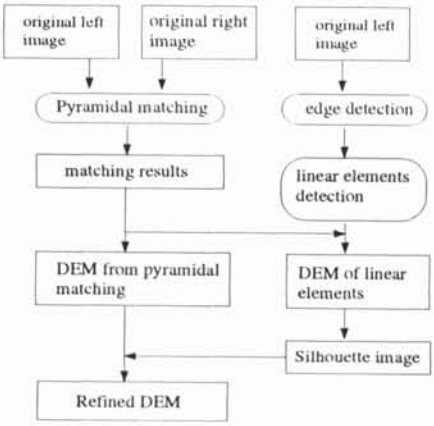

Figure 1: Summary of the overall procedure

RESULTS

Test images are shown in Fig 2. They were digitised from two aerial photographs with pixel size of 15.42 μm and 12.80 μm in x and y direction respectively. The size of test images are 1640 pixels by 1218 pixels and the resolution is about 15 cm.

A 6 level image pyramid was generated by reducing the image by a factor of 2 and pyramidal matching was applied with automatic seed points and tile-based seed selection strategy. The "Gotcha" with 69 manually selected seed points was also applied to the test images for comparison. Seed points were selected to cover as many isolated regions as possible. DEMs were created from both methods and compared with the reference DEM generated from photogrammetric measurements. Table 1 summarises the error statistics of DEMs and Fig 3 shows DEMs. As shown in table 1 and Fig 3, the pyramidal matching algorithm proposed in this paper enables an automated 3D coordinates extraction system. However, the reliability and accuracy of the DEM from pyramidal matching are poorer than those from the "Gotcha", which are to be improved by the use of a silhouette mask.

Figure 2: Left and right images

Table 1: statistics of elevation errors of DEMs

	DEM from Gotcha	DEM from pyramid	refined DEM
coverage [a]	13362	16354	14618
accuracy [b] (m)	4.709	5.114	3.934
reliability [c]	4.22%	4.62%	3.66%
min.(m)	-25.89	-26.36	-29.36
max.(m)	29.93	29.67	29.67
average(m)	0.192	0.515	-0.190
St. dev(m)	4.705	5.088	3.929

[a] Coverage of a DEM is defined by the number of points in a DEM.

[b] Accuracy is defined by the root mean square elevation error.

[c] Reliability is defined as the proportion of points with elevation errors greater in magnitude than 3σ, where σ is standard deviation of errors.

Edge elements were detected by the CPK filter of the size of 7 pixel × 7 pixels. A Hough transformation and a connected edge labelling algorithm were applied to the edge elements and Linear elements were obtained by combining results from both methods. A DEM was generated by assigning height to a linear element according to the pyramidal matching results and it was used to create a silhouette image of buildings. A silhouette image is shown in Fig 4.

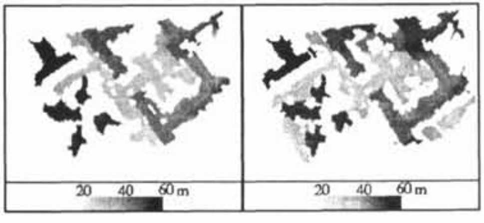

Figure 3: Left: DEM from Gotcha. Right: DEM from pyramidal matching(6 levels)

The silhouette image was used to refine the DEM from pyramidal matching. The results are summarised in table 1. Fig 4 shows the DEM refined by the silhouette mask. The statistics show that the operations used to remove potential blunders around boundaries of buildings were successful. The refined DEM has better coverage, accuracy, and reliability compared with the DEM from the "Gotcha".

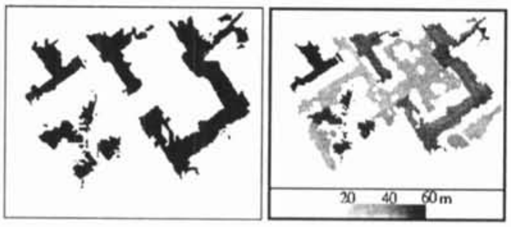

Figure 4: Left: Silhouette image. Right: DEM refined by the silhouette mask

AUTOMATIC BUILDING DETECTION

A monoscopic approach for building detection was developed as a counterpart of the stereoscopic algorithm in the earlier section for more complete urban area image understanding. It is based on the traversal of a graph constructed from lines and their relations.

Edge elements are detected by the CPK filter as before. Due to the relative simplicity of finding the length and end points of a line compared to Hough transformation, a connected edge labelling algorithm is used for line extraction. The absolute difference of angle between neighbouring pixels is examined to select linearly connected edge elements. Two end points of a linear elements are searched with end point templates, and a line is defined with them. Lines undergo a merging process to combine small segments broken from a long line and remove closely located parallel lines.

After defining the start and finish points from two end points of a line, the relations between the lines are classified by *positive connection*, *negative connection*, *neutral connection* , and *parallel connection*. Positive connection is when a line is connected with another line near the finish point. Negative connection and neutral connection are defined similarly with the start point and the centre point. A line is divided into three parts of equal length to define positive, negative, and neutral connections. To reduce the size of the graph and the computation time, relations of lines which are shorter than a threshold and whose relative angle is not near vertical are not considered.

Line relations are stored in a graph. Nodes of the graph consist of lines. Relations between lines are represented by arcs between the nodes in the graph.

Building hypotheses are generated by finding closed loops in line relation graph. A depth-first search algorithm is used for the traversal of graph. In order to make sensible building hypotheses, the search is limited by a rule that the type of connection of the current line of a search chain with respect to the next line should be different from that of the current line with respect to the previous line. The search is also limited by the number of nodes it visits.

Building hypotheses are examined to select valid building hypotheses. First of all, similar building hypotheses are merged. Similarity is defined when building hypotheses have all identical components except one component in their search chain. False building hypotheses made from the ground level lines are then removed after detecting vertical shadow lines and vertical lines. Vertical shadow lines, which refer to the shadow lines from the vertical side of buildings, are detected by using the fact that their orientation aligns with the direction of illumination. Vertical lines are detected as they are directed toward the vertical vanishing point [10]. The ground level lines are determined as they are connected with vertical shadow lines in the opposite direction to the Sun and with vertical lines in the direction to the focal point. The direction of illumination and the position of focal point relative to image coordinates are assumed to be known. By examining the length of vertical lines, the height of verified building hypotheses are also estimated.

RESULTS

The left image in Fig 2 was tested with the building extraction algorithm. 2313 lines were found after applying edge detection and connected edge labelling algorithm. 1468 lines were selected after merging and they are shown in the left image of Fig 5. Lines from buildings in the lower left part of image are not detected well due to the complicated roof structures and heavy shadows and this is the limitation of our current line detection system. 2334 line relations were found with a length threshold of 30 pixels and an angle threshold of 60^o. A graph with 1468 nodes and 2334 arcs was created and 2166 building hypotheses were found by limiting to a maximum 6 visits. The right image in Fig 5 shows building hypotheses.

Figure 5: Left: 1468 Lines. Right: 2166 Building hypotheses

The merging process was applied to building hypotheses(BH) and 472 building hypotheses were selected. They are shown in the left image of Fig 6. 36 vertical shadow lines and 104 vertical lines were detected and the false building hypotheses were removed. 384 building hypotheses were finally selected. They are shown in the right image of Fig 6 overlapped with the original image. Comparison with the ground truth buildings tells that 96.2% of ground truth buildings were recovered but 10.9% of buildings detected by our building detection system were proved erroneous. The height of the buildings was estimated by vertical lines and Fig 7 shows the perspective view of the buildings.

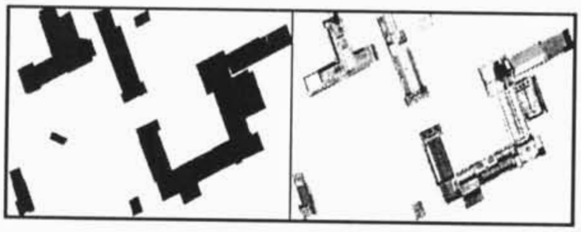

Figure 6: Left: BH after merging. Right:Final BH

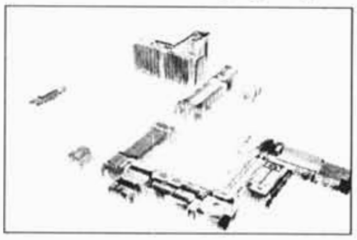

Figure 7: Perspective view of buildings

A 2 m resolution Russian space image(DD5) was also tested. Fig 8 shows a DD5 image over Berlin and a line image extracted from the DD5. 437 lines were extracted. However, lines from small buildings were not extracted due to the resolution of image. To avoid the ground level lines being included in building hypotheses, a threshold value of 75^o was used for angles between lines. 35 building hypotheses were generated and 10 building hypotheses were verified after merging similar building hypotheses. They are shown in Fig 8 overlapped with the original image. Detecting vertical shadow lines and vertical lines were not performed as they are not visible.

CONCLUSIONS

In this paper, two approaches for urban area image understanding are introduced. For a stereoscopic approach, an automated pyramidal matching algorithm with the use of automatic seed points is proposed. It is also emphasised that an appropriate control strategy is necessary for a pyramidal approach, and a tile-based selection strategy is adopted. Removal of possible blunders is performed by the use of linear elements from buildings. The first quantitative assessment of accuracy for the building height extraction algorithm over urban areas is performed. Height of buildings were extracted with a RMS. error of 4 m.

A novel monoscopic building detection algorithm is also proposed as a counterpart of the stereocopic approach. It stores lines and line relations in a graph and finds building hypotheses by a depth-first graph traversal algorithm. Building hypotheses are verified by merging and removing false building hypotheses. 15cm resolution digitised aerial photograph and 2m resolution space-borne image were tested with the new algorithm. Complicated buildings in aerial photograph and large buildings in space-borne image are well detected.

Figure 8: Left: A DD5 image over Berlin airport. Middle: Lines. Right: Building hypotheses

ACKNOWLEDGEMENTS

We would like to thank WIB for the use of the DD5 image over Berlin.

REFERENCES

[1] J-P Muller, "Real-time Stereo Matching and its role in future mapping systems", *Proc. Surveying and Mapping 89, University of Warwick*, **17-21**, April 1989

[2] J-P. Muller, A. Anthony, A.T. Brown, A.T. Deacon, S.A. Kennedy, P.M. Montgomery, G.W. Robertson, and D.M. Watson, "Real-time stereo matching using transputer arrays for close-range applications", *Proc. joint IAPR Workshop on "Computer Vision - Special Hardware and Industrial Applications" held at Nihon Daigaku, Tokyo, Japan*, **45-49**, October 12-14, 1988

[3] M.A. O'Neill and M.I. Denos, "Practical Approach to the stereo matching of urban imagery", *Image and Vision Computing*, **10(2):89-98**, March 1992

[4] A.W. Gruen, "Adaptive Least Squares Correlation: A Powerful Image Matching Technique", *S. Afr. Journal of Photogrammetry, Remote Sensing and Cartography*, **14(3):175-187**, 1985

[5] G.P. Otto and T.K. Chau, "A Region-Growing Algorithm for Matching of Terrain Images", *Image and Vision Computing*, **7(2):83-94**, 1988

[6] T. Kim and J-P. Muller, "Automated Urban Area Building Extraction from High Resolution Stereo Imagery", *submitted to Image and Vision Computing*, 1994

[7] J. Canny, "A computational Approach to Edge Detection", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **PAMI-8(6):679-697**, Nov. 1986

[8] M. Petrou and J. Kittler, "Optimal Edge Detectors for Ramp Edges", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **PAMI-13(5):483-491**, 1991

[9] R.O. Duda and P.E. Hart, "Use of the Hough Transformation to Detect Lines and Curves in Pictures", *Commum. ACM*, **15:11-15**, Jan. 1972

[10] M. Herman and T. Kanade, "The 3D Mosaic Scene Understanding System", *From Pixels to Predicates edited by A.P. Pentland*, **322-358**, 1986