RECONSTRUCTING 3D BUILDING WIREFRAMES FROM MULTIPLE IMAGES

Ahmed F. Elaksher, James S. Bethel, and Edward M. Mikhail

School of Civil Engineering, Purdue University, 1284 Civil Engineering Building, West Lafayette, IN 47906, USA
elaksher@ecn.purdue.edu bethel@ecn.purdue.edu mikhail@ecn.purdue.edu

ABSTRACT

Building extraction in urban areas is one of the difficult problems in image understanding and photogrammetry. Building delineations are needed in cartographic analysis, urban area planning, and visualization. Although one pair of images is adequate to find the 3D position of two visibly corresponding image features it is not sufficient to extract the entire building due to hidden features that are not projected into the image pair. This paper presents a new technique to detect and delineate buildings with complex rooftops by extracting roof polygons and matching them using multiple images.

The algorithm discussed in this paper starts by segmenting the images into regions. Regions are then classified into roof regions and non-roof regions using a two-layered Neural Network. A rule-based system is then used to convert the roof boundaries to polygons. Polygon correspondence is established geometrically, all possible polygon correspondent sets are considered and the optimal set is selected. Polygon vertices are then refined using the known geometric properties of urban buildings to generate the building wire-frames. The algorithm is tested on a number of buildings and the results are evaluated. The RMS error for the extracted building vertices is 0.25m using 1:4000 scale aerial photographs. The results show the completeness and accuracy that this method can provide for extracting complex urban buildings.

1. INTRODUCTION

Recent research in the area of building extraction covers building extraction from aerial images, digital elevation models (DEM), thematic maps, and terrestrial images. Aerial images and digital elevation models are the primary data sets used in most building extraction systems. Some systems use only aerial images; some use DEM only; others use both data sets. In (Suveg and Vosselman, 2000) thematic maps and GIS databases are used to help resolving ambiguity in the extracted buildings or in generating building cues that can be refined by DEM, aerial images, or both. In (Chein and Hsu, 2000) one pair of images is used to extract buildings. This is insufficient since parts of the buildings can be either obscured by other features or not projected into this specific pair. Although in (Kim and Nevatia, 1999) more than one pair of images is used to extract the buildings, the matching was carried out in a pairwise fashion.

In (Brunn and Weidner, 1997) the DEM is used solely to extract the building models. Researchers using only DEM in building extraction start by segmenting the DEM. This process is problematic with image-driven DEM and LIDAR-driven DEM. Outliers often disturb the extracted regions. Some researchers compute the slope of the model surface in both directions to provide more information that can assist the building extraction. However slopes are not always accurate due to outliers in the digital elevation models. In (Wang, 2000) a building extraction system from LIDAR data is presented. His results suffer from some of the same problems that occur with image-driven DEM.

In (Zhao and Trinder, 2000) and (Seresht and Azizi, 2000) aerial images and DEM are used for the building extraction process. They started with the DEM to provide building regions or building cues and then they used the images to refine the extracted building regions.

In (Fischer *et. al*., 1999) and (Förstner, 1999) the building extraction problem was solved using a semi-automated approach. The user has to define the building model and find the building elements in one image by a number of mouse clicks. Then the algorithm finds the corresponding features in other images, and matches them to build the 3D wire-frame of the building. This approach supports the extraction of more complex buildings; however it requires the user to spend a great amount of time interacting with the system.

In this article a new technique to extract urban area building wire-frames using more than one pair of images is presented. The input to the algorithm is a number of images for the building. The minimum number required is two; however this number is generally not enough for a complete extraction. In this research four aerial images per building are used in the extraction process.

We start by segmenting the images using a split and merge image segmentation technique. The extracted regions are then classified into roof regions and non-roof regions using a two-layered Neural Network. Two attributes are used in the classification process. The first attribute measures the linearity of the building borders. The second attribute measures the average elevation of the region and it is driven from a digital elevation model. The borderlines for the building regions are then extracted from the border pixels using a modified version of the Hough transformation. A rule-based system is then employed to convert the extracted lines to polygons. The algorithm can extract either triangle or quadrilateral roof facets. Correspondence between roof polygons is established using the geometrical properties of the polygons. A least squares estimation model is implemented to find corresponding polygons. Geometric constrains between vertices in one polygon, symmetric planes, and horizontal planes are utilized in the least squares model to refine the extracted building vertices.

The algorithm has been tested on a large sample of buildings selected quasi-randomly from the Purdue University campus. Four images are used for each building and the automatically extracted wire-frames for the extracted buildings are

presented. The RMS error between manually extracted coordinates and the produced coordinates for six buildings is 0.25 meter and only two vertices were missing. These results suggest the completeness and accuracy that this method provides for extracting complex urban buildings.

Section 2 explains the split and merge image segmentation technique. In section 3 the region classification process is discussed. Section 4 presents the region to polygon conversion. The multi image 3D polygon extraction algorithm is explained in section 5. Results are given in section 6. Conclusions are discussed in section 7.

2. IMAGE REGION EXTRACTION

In this section the process of extracting image regions is presented. Image segmentation could be done using a wide range of techniques. The best technique we have found for segmenting aerial images is the split and merge image segmentation technique. The split and merge image segmentation technique has three main steps. First splitting the image: the image is recursively divided into smaller regions until a homogeneity condition is satisfied. Then adjacent regions are merged to form larger regions based on a similar criterion. In the last step, small regions are either eliminated or merged with larger regions. The criterion used in the split and merge image segmentation method is that the difference between the minimum and maximum intensities in any region is less than a certain threshold. More details can be found in (Horowitz and Pavlidis, 1974) and (Samet, 1982). The results of the split and merge image segmentation technique for five sample buildings, are shown in Figure 1-a, b, c, and d.

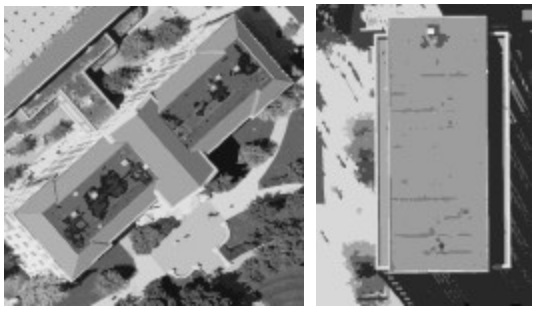


Figure1-a and b. Split and Merge Image Segmentation Results for 2 Buildings

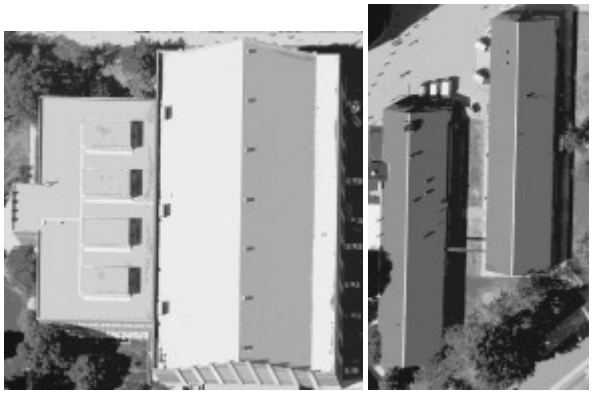


Figure1-c and d. Split and Merge Image Segmentation Results for 3 Buildings

3. REGION CLASSIFICATION USING NEURAL NETWORKS

A Neural Network is implemented to distinguish roof regions from non-roof regions. Each region is assigned two attributes for the classification process. The first attribute measures the linearity of the region boundaries, while the second attribute measures the percentage of the points in the region that are above a certain height.

3.1. Region Border Linearity Measurement

After segmenting the building images a modified version of the Hough transformation is employed to measure border linearity. The approach includes the following steps; extracting region border points, linking border points, finding local lines that fit groups of successive points, and filling a parameter space similar to the Hough parameter space for line extraction. The parameter space is then searched and analyzed to determine a measure for the border linearity, (*BL*), Equation 1. The border linearity is measured as the percentage of the sum of the number of points in the larger four cells in the parameter space to the total number of border points. Figure 2-a shows a parameter space for a roof region, while Figure 2-b shows a parameter space for a non-roof region.

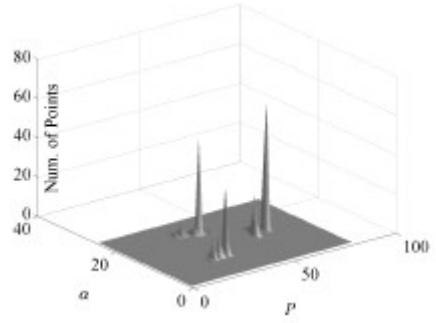


Figure 2-a. The Modified Hough Parameter Space for the Border of a Roof Region

$$BL = \frac{\text{Number of Points in Larger 4 Cells}}{\text{Total Number of Border Points}} \quad (1)$$

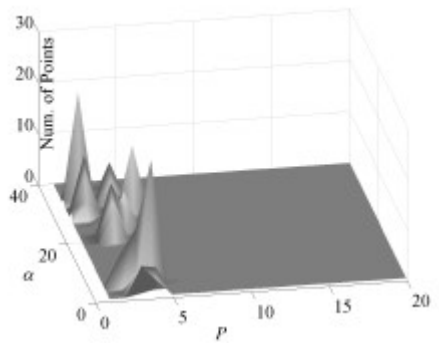


Figure 2-b. The Modified Hough Parameter Space for the Border of a Non-Roof Region

3.2. Region Elevation Measurement

The second attribute assigned to the roof regions quantitizes the height of each region (*RH*), Equation 2. A digital elevation model is used for this task. First each point in the image is assigned an elevation value by projecting the DEM back to the image using the image registration information, the pixel location in the image, and the DEM. For each image point a ray is generated starting from the exposure station of the camera and is directed toward the point. The intersection between the ray and the DEM defines the elevation of the image point. The *RH* is measured as the percentage of the number of the roof region points that are above a certain elevation to the total number of points in the region.

$$RH = \frac{\text{Number of Region Points Above } H_{min}}{\text{Total Number of Region Points}} \qquad (2)$$

Where H_{min} = Min Building Elevation

3.3. Implementing the Neural Network

Figure 3 shows a 2D plot for the two region attributes, the total number of regions is 2081 regions, 623 regions are roof regions and the rest are non-roof regions. A simple two-layered Neural Network is used to discriminate between roof and non-roof regions, Figure 4. The activation function for all nodes is the Sigmoid Function, (Principe *et. al.*, 1999).

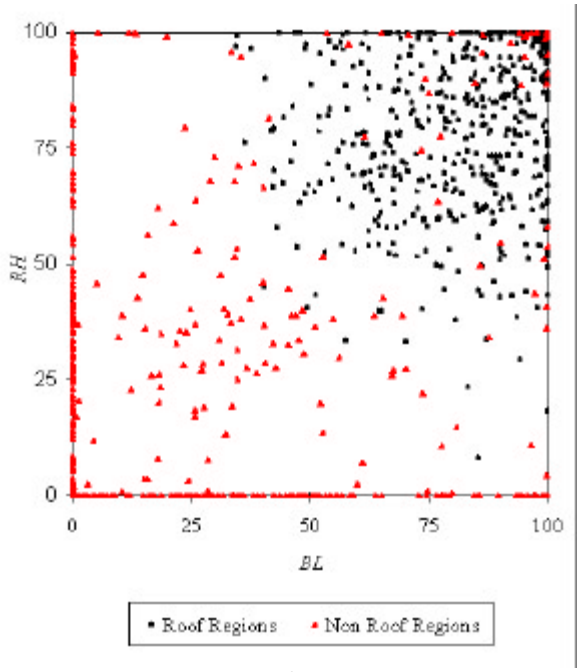


Figure 3. Scatter Diagram of Border Linearity (*BL*) vs. Region Height (*RH*)

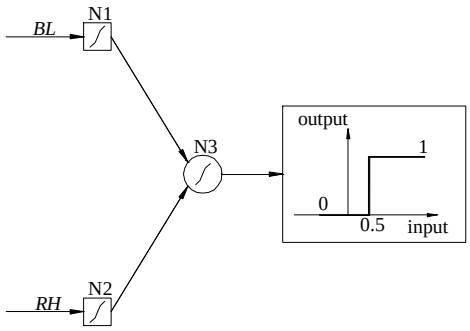
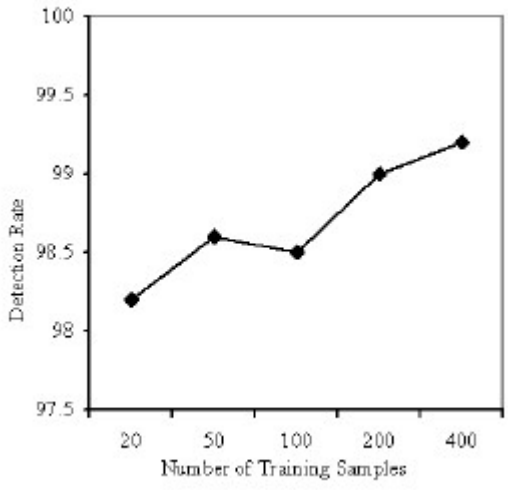


Figure 4. The Implemented Two-layered Neural Network

To study the performance of the Neural Network a variety of training data sets were used with different sizes. The training data set sizes used are 20, 50, 100, 200, and 400 samples. For each training data set size the experiment was performed 10 times using a non-overlapping randomly selected training data set. The average detection rate and false alarm rate for each training data set size is recorded and shown in Figures 5-and b.



Figur5-a. The Detection Rate vs. Training Data Set Size

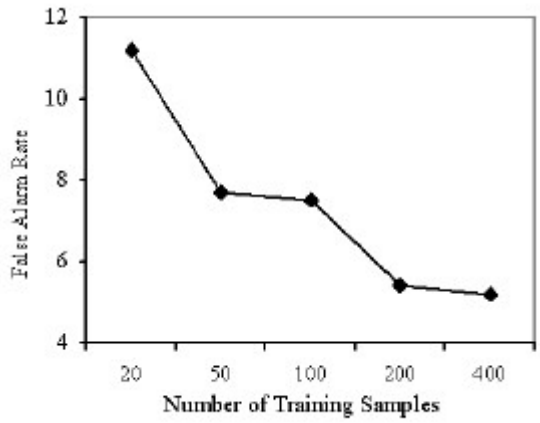


Figure 5-b. The False Alarm Rate vs. Training Data Set Size

Increasing the training data set size doesn't affect the detection rate significantly, we can see that the range of the detection rate is between 98.2% and 99.2%; we gain only 1% in the detection rate when we increase the size of the training data set from 20 samples to 400 samples. However increasing the size of the training data set has a significant effect on the false alarm rate. Increasing the size of the data set from 20 samples to 400 samples reduce the false alarm rate from 11.5% to 5.5%; we gain 6.0% improvement in the false alarm rate by increasing the size of the training data set from 20 samples to 400 samples. The results using 100 training samples are used in the rest of this research.

4. CONVERTING REGIONS TO POLYGONS

The 2D modified Hough space discussed in the previous section is helpful in extracting the border lines for the roof regions. Given all points contributing to a certain cell, a nonlinear least squares estimation model is used to adjust the line parameters given the cell locations in the parameter space as approximate values for the line represented by this cell. Lines are then grouped recursively until no more lines with similar parameters are left. Short lines are then rejected. Figure 6 shows the extracted border lines for two buildings.

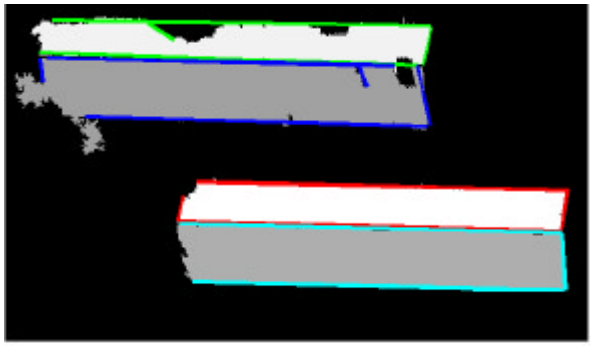


Figure 6. Extracted Border Lines for 2 Buildings

The next step is to convert the extracted lines to polygons using a rule-based system. The rules are designed as complex as possible to cover a wide range of polygons. Figure 6 shows the challenging in converting the extracted border lines to polygons. For some polygons they might be quadrilaterals, however only three borderlines are detected. Some quadrilateral regions might have more than four borderlines detected. The mechanism that is developed in this research works in three steps. The first step is to find all the possible intersections between the borderlines. However if the two lines are almost parallel the intersection point is not considered. If the distance between the end point of the line and the intersection point is large the intersection point is rejected. The next step is to generate a number of polygons from all the recorded intersections. Each combination of four or three intersection points is considered to be a polygon hypothesis. Some hypotheses are ignored if the difference in area between the region and the hypothesized polygon is more than 50%. If the internal angles between the intersected lines is out of the range [30^{o}-150^{o}] the hypothesis is discarded. The third step is to find the optimal polygon that represents the region borders. The best polygon that represents the region is chosen from the remaining polygons. A template matching technique is used to find the best polygon that represents the region. Figure 7 shows the extracted polygons for two buildings.

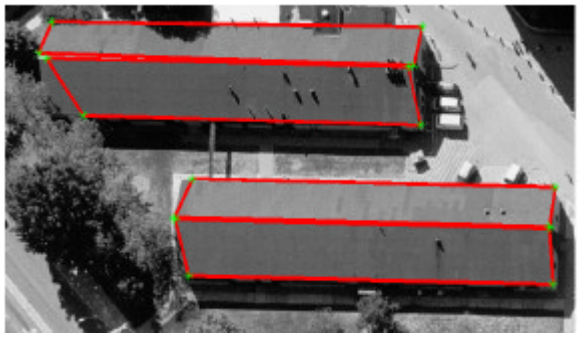


Figure 7. Extracted Image Polygons for 2 Buildings

5. 3D POLYGON EXTRACTION

In this section the process of finding the correspondence polygons among all images and matching them is discussed.

5.1. Polygon Correspondence

After finding the building roof polygons in the images, we start finding the correspondence polygons. We designed a new technique to find correspondence polygons based on their geometrical properties. All possible polygon correspondence combinations are considered and for each combination the vertices of the correspondent polygons are matched across all available views, since we have more than one pair of images we can calculate the residuals of the matched 3D polygon vertices. The matching residuals are summed for each combination set, and the combination set with the minimum residual is selected as the best combination set. Figure 8 describes the process of finding the correspondence polygons in four images. In order to minimize the running time, some subsets are rejected before the matching process using the epipolar geometry and the minimum and maximum building heights.

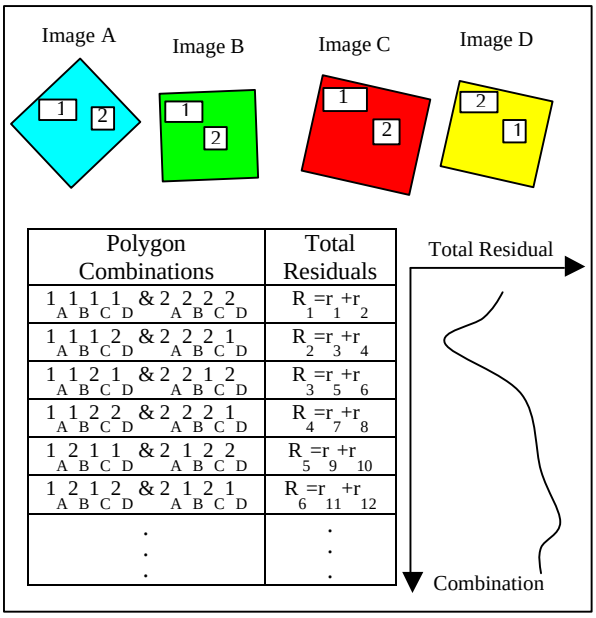


Figure 8. The Multi Image Polygon Matching Process

5.2. 3D Polygon Coordinates Refinement

After finding the corresponding polygons from the previous step, the 3D coordinates for each roof polygon is computed, however the building topology is not yet reconstructed. We implement a geometrically constrained least squares model in order to refine the locations of the polygon vertices and to reconstruct the building topology. The input observations are the image coordinates of the polygon vertices, the unknowns are the object space coordinates for the 3D polygons, however we have to take into consideration the following constraints:
1-The polygon vertices should be in the same plane.
2-Symmetric polygons should be constrained to have symmetric parameters.
3-Points that are almost in a horizontal plane are constrained to have the same elevation.
4-Nearby vertices should be grouped into one vertex.

The aim of the refining step is to convert groups of neighboring vertices into one vertex, adjust the elevations of horizontal points, and reconstruct the correct relativity relation between adjacent facets.

6. RESULTS

In the following section the results of extracting the 3D building wire-frames are shown. Figure 9 shows a sample of 17 buildings extracted using the presented algorithm. The results show the completeness and accuracy of the 3D roofs that can be extracted using this system.

In order to evaluate the accuracy of the extracted buildings, the 3D coordinates of 6 building vertices were extracted manually and compared with the automatically extracted ones. The RMS error for the vertices in all six buildings is 0.25m. Table 1 shows the detailed analysis for the evaluated 6 buildings. Seventy-eight vertices were detected out of 80 in the 6 buildings.

Building	(X,Y) RMS	(Z) RMS	Missing Vertices
BLD 1	0.22	0.12	0
BLD 2	0.32	0.24	1
BLD 3	0.22	0.37	1
BLD 4	0.42	0.24	0
BLD 5	0.22	0.25	0
BLD 6	0.22	0.27	0

Table 1. Results for Extracting Six Buildings Roofs, RMS in meters

7. CONCLUSIONS

The results presented in this paper show the great improvement that this algorithm adds to the current building extraction techniques. The algorithm succeeds in extracting a wide range of urban building. The tested data set includes simple buildings with one rectangular roof, gabled roof buildings, multi store buildings with large relief, and a variety of complex buildings.

The RMS error is about 0.25m. The false regions that were wrongly classified in the Neural Network were automatically eliminated since they didn't have any correspondence. The overall detection rate for both the Neural Network classification and the 3D reconstruction is 97.5%. The algorithm succeeded in matching the image polygons simultaneously across more than two images, this reduced the false alarm matches and increased the result accuracy. The method can be implemented using any number of images. More work is necessary and will be carried out in the future to improve the building delineations even further.

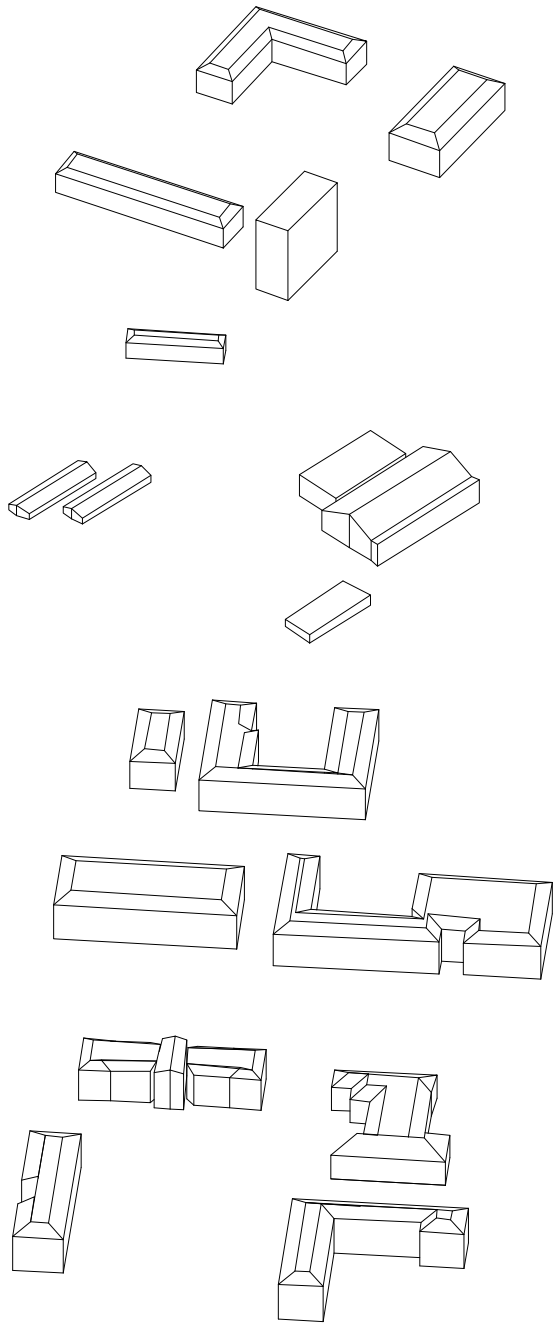


Figure 9. The Wire-Frames of a Sample of the Extracted Buildings

REFERENCES

Brunn A. and Weidner U., 1997. Extracting Buildings from Digital Surface Models. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Stuttgart, Germany, Vol. 32, Part 3-4W2, pp. 27-34.

Chein L. and Hsu W., 2000. Extraction of Man-Made Buildings in Multispectral Stereoscopic Images. In: *Proceeding of the 19th ISPRS Congress*, Amsterdam, The Netherlands, Vol. XXXIII, Part B3/1, pp. 169-176.

Fischer *et. al.*, 1999: Fischer A., Kolbe T.H., and Lang F., 1999. On the Use of Geometric and Semantic Models for Component-Based Building Reconstruction Proceedings. In: *Semantic Modeling for the Acquisition of Topographic Information from Images and Maps Workshop*, München, Germany, pp. 101-119.

Förstner W., 1999. 3D-City Models: Automatic and Semiautomatic Acquisition Methods. In: *Photogrammetric Week*, Stuttgart, Germany, pp.291-303.

Horowitz S.L. and Pavlidis T., 1974. Picture Segmentation by a Direct Split and Merge Procedure. In: *Proceeding of 2nd International Conference on Pattern Recognition*, Copenhagen, Denmark, pp. 424:433.

Kim Z. and Nevatia R., 1999. Uncertain Reasoning and Learning for Feature Grouping. *Computer Vision and Image Understanding*, Vol. 76, No. 3, pp. 278-288.

Principe *et. al.*, 1999: Principe J., Euliano N., and Lefebvre W., 1999. *Neural and Adaptive Systems: Fundamentals Through Simulation*, John Wiley and Sons, NY 10158-0012.

Samet H., 1982. Neighbor Finding Techniques for Images Represented by Quadtrees. *Computer Graphics and Image Processing*, Vol. 18, No.1, pp. 37-57.

Seresht M. and Azizi A., 2000. Automatic Building Recognition from Digital Aerial Images. In: *Proceeding of the 19th ISPRS Congress*, Amsterdam, The Netherlands, Vol. XXXIII, Part B3/2, pp. 792-798.

Suveg I. and Vosselman G., 2000. 3D Reconstruction of Building Models. In: *Proceeding of the 19th ISPRS Congress*, Amsterdam, The Netherlands, Vol. XXXIII, Part B2, pp. 538-545.

Wang Z., 2000. Building Extraction and Reconstruction From LIDAR Data. In: *Proceeding of the 19th ISPRS Congress*, Amsterdam, The Netherlands, Vol. XXXIII, Part B3/2, pp. 958-964.

Zhao B. and Trinder J., 2000. Integrated Approach Based Automatic Building Extraction. In: *Proceeding of the 19th ISPRS Congress*, Amsterdam, The Netherlands, Vol. XXXIII, Part B3/2, pp. 1026-1032.

ACKNOWLEDGMENT

The authors would like to thank the Army Research Office for sponsoring this work.