# Automatic Building Detection in Aerial and Satellite Images

Parvaneh Saeedi
School of Engineering Science
Simon Fraser University
Burnaby, BC, Canada
psaeedi@sfu.ca

Harold Zwick
Department of Research and Development
MacDonald, Dettwiler and Associates Ltd.
Richmond, BC, Canada
hhz@mdacorporation.com

*Abstract*—Automatic creation of 3D urban city maps could be an innovative way for providing geometric data for varieties of applications such as civilian emergency situations, natural disaster management, military situations, and urban planning. Reliable and consistent extraction of quantitative information from remotely sensed imagery is crucial to the success of any of the above applications. This paper describes the development of an automated roof detection system from single monocular electro-optic satellite imagery. The system employs a fresh approach in which each input image is segmented at several levels. The border line definition of such segments combined with line segments detected on the original image are used to generate a set of quadrilateral rooftop hypotheses. For each hypothesis a probability score is computed that represents the evidence of true building according to the image gradient field and line segment definitions. The presented results demonstrate that the system is capable of detecting small gabled residential rooftops with variant light reflection properties with high positional accuracies.

*Index Terms*—Building extraction, satellite image processing, aerial image processing, photogrammetry, computer vision, geometrical shape extraction.

## I. INTRODUCTION

The problem of detection and characterization of 3D buildings in the urban/suburban areas is a very complicated one that has many applications in a vast variety of areas. Automatic creation of 3D urban city maps could be an innovative way for providing geometric data for varieties of applications such as civilian emergency situations, natural disaster management (flooding, earthquakes, and landslides), military situations (active engagement of force, counter terrorism and peace keeping measures), urban planning, airport hazard analysis, and statistical geographic localization (such as health, crime, and past natural disasters). Reliable and consistent extraction of quantitative information from remotely sensed imagery is crucial to the success of any of the above applications.

Almost all operational approaches developed over the years for 3D building/map reconstruction are semi-automated ones, where a skilled human operator is involved in the 3D geometry modeling of building instances. Perhaps the most key role of the human operator in such systems is the identification of the building rooftops by, for instance, drawing lines and curves that depict buildings walls and borders. This is an expensive and tedious process with a low update rate.

While numerous semi-automated systems have been developed, a limited number of automated systems are reported in the literature. Some of these works present instances of limited good results especially for aerial imageries (higher resolution) and larger size buildings. These systems are still far from being capable of coping with the existing complexities of urban structures and maps. A system that can robustly detect and identify/measure building structures with high reliability is yet to be developed.

### A. Previous work

Prior building extraction systems generally consider two main processes [2]: footprint detection and 3D reconstruction. In this concept, building outlines and roof structures can be described with the use of lines, regions, planar patches, polyhedral shapes, geometrical models, and multiple images [8], [17], [21], [7].

Early works in the area of building detection are mostly based on the main assumption that the buildings have quadrilateral image footprints and therefore the detection model has a quadrilateral shape. With such assumption, almost all the previous works have used edge/line detection techniques (Canny, DoG, Laplace, Hough, Boldt, Weiss, gradient values/directions) [15], [12], [10], [5], [13] to extract straight lines. The relationships between extracted lines are identified and classified to generate building hypotheses. This group of works generally has shown some success when detecting large buildings with flat rooftops. They however tend to fail (or suffer instabilities) in cluttered scene scenarios or scenes including small buildings.

Some approaches utilize image cues and geometric constraints with sophisticated methodologies to create more complex primitives. These primitives are then matched against predefined models. Cheng et al [4] introduces a segmentation method based on 2D histograms for partitioning the aerial images into four distinctive regions. [16] uses [4]'s image primitives to detect buildings by employing a modified partial snake model. [9], [11] propose sophisticated surface fitting methods using image corners. [20] devises a system for detecting flat or gabled roofs using textured segments and image corner points that are matched at the pixel level using mutual information. Sohn and Dowman [19] suggest an automatic building extraction technique using local Fourier analysis to determine the dominant orientation angle of a

building cluster in dense urban areas of IKONOS imagery. The main problem with this group of work is that only a small number of overall features is used for building identification, and therefore these methods usually have a low true-positive rate.

### B. Objective

In this work the problem of automatic building rooftop detection, as a key element of the 3D map reconstruction, is addressed. The system is designed with the following assumptions:

- The images to be processed are panchromatic: This is to restrict the scope of this work to the fundamental problem.
- Space-borne imagery: To develop an algorithm that find smaller buildings on lower resolution imagery.
- The walls are vertical: This assumption applies within each rectangular building component.
- The imagery is acquired on a clear day, with the sun not on the horizon: This will ensure that good shadows may be available as supporting evidences of rooftop structure.
- Quadrilateral footprints: The building rooftops could be closely fitted within a quadrilateral shape.

The primary novelty of this work is a methodology for detecting building outlines with an accuracy within a pixel on general satellite/aerial imageries (some obtained from Google Maps). Moreover the system is capable of processing smaller suburban buildings with gabled rooftops.

### C. Paper outline

The basis of 2D image primitives detection is explained in Section II. Section III describes details for hypothesis generation, validation and refinement. Section IV represents the experimental results. Conclusions and future work are represented in Section V.

## II. 2D IMAGE PRIMITIVES

The goal of this section is to apply a series of image processing algorithms to an input image to extract image primitives that are used in later stages. The outputs of this section are:

- The gradient field of the image, including both gradient magnitude and direction.
- A set of straight line segments (location, extent, orientation) found in the image.
- A set of image segments that represent similar image regions at various sensitivities.

### A. Gradient field

The objective of this step is to generate the gradient field of the image including gradient magnitude and direction at each image pixel. For this purpose first a $3\times3$ pixel Gaussian low-pass filter is applied over the image to reduce pixel-level noise. The horizontal and vertical gradients are then computed at each pixel by convolving $2\times2$ pixel horizontal and vertical masks across the image. The gradient magnitude and direction are estimated from the horizontal and vertical gradients at each pixel.

### B. Straight line extraction

The objective of this step is to extract a set of straight-line segments from the image, given the gradient field computed in Section II-A. The algorithm implemented to achieve this goal is the Burns line detector [3], which utilizes both the gradient magnitude and gradient orientation to form line support regions and eventually straight line segments. The following steps describe this procedure.

1. Partition pixels into bins based on the gradient orientation values. A bin size of 45 degrees was selected.
2. Run a connected-components algorithm to form line support regions from groups of 4-connected pixels that share the same gradient orientation bin.
3. Eliminate line support regions that have an area smaller than a specified threshold.
4. Repeat steps 1, 2, and 3 by shifting the gradient bins to produce a second set of line support regions. This accounts for the possibility that some true lines may have pixels that lie on either side of an arbitrary gradient orientation boundary.
5. Use a voting scheme to select preferred lines from the two sets (original and shifted) of candidate lines.
6. For each line support region, compute the line represented by that region by performing a least squares fit.

### C. Line linking and filtering

The objective of this step is to link collinear line segments that are separated by very small gaps. Following algorithm describes linking process:

1. Sort the lines in the order they would be encountered if a horizontal sweep was performed across the image.
2. Use a divide-and-conquer method to efficiently determine nearby pairs of lines.
3. Test each pair of nearby lines to determine whether they should be linked. All 5 criteria, which are illustrated in Figure 1, must be satisfied for a pair of lines to be linked.

Two line filtering processes are applied to remove lines that are not likely to positively contribute to rooftop hypothesis generation process.

- The first filter acts on the line length. In general, long lines are more likely to contribute to hypothesis formation and subsequent steps.
- The second filter employs the gradient across the line segment to separate regions of high contrast from low contrast. In this process the gradient across lines are normalized with respect to the average intensity across the scene.

All detected lines at this stage are saved for future use in Section III. Figure 2 depicts the lines for the sample imagery after the linking and filtering processes.
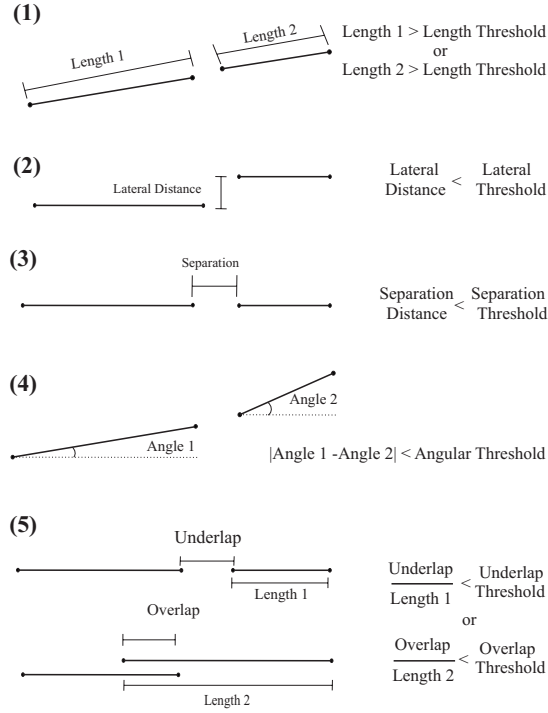
Fig. 1. Five required criteria for line linking process.



Fig. 2. Detected lines after liking and filtering processes.

## D. Image segmentation

The objective of this process is to segment input image into connected regions that fully or partially overlap with areas corresponding to the rooftops.

For this purpose a graph-based segmentation [6] is employed that is capable of preserving details in low-variability image regions while ignoring details in high-variability areas. The algorithm includes following steps:

1. Each image pixel is considered as a region where it corresponds to a node ($v \in V$) in the overall image graph of $G(V, E)$.
2. Neighboring pixels are connected by undirected edges ($e \in E$). For each edge a weight coefficient is computed according to the dissimilarities between pixels.
3. Similar regions $A$ and $B$ are merged together to produce a larger region if the following condition is held:

$$Dif(A, B) \leq MInt(A, B) \tag{1}$$

where

$$Dif(A, B) = \min_{v_i \in A, v_j \in B, (v_i, v_j) \in E} w((v_i, v_j)) \tag{2}$$

Here $E$ is the graph edge set and $w((v_i, v_j))$ is the weight between vertex $v_i$ and $v_j$.

$$MInt(A, B) = \\ \min \ (Int(A) + \tau(A), Int(B) + \tau(B)) \tag{3}$$

$$Int(A) = \max_{e \in MST(A, E)} w(e) \tag{4}$$

MST represents the Minimum Spanning Tree graph $G(V, E)$.

$$\tau(A) = \frac{k}{|A|} \tag{5}$$

The value of $k$ defines the sensitivity level. Since each scene could include a number of buildings with different sizes, each input image is segmented with eight different values of $k$. All the segments generated from various $k$s are placed in an image segment dataset.

In order to improve segmentation results, small modification is applied to the above algorithm by replacing the calculation of the Euclidean color distance with a closer representation to the human brain's perception [1] using following equations:

$$\Delta C = \\ \sqrt{(2 + \frac{\overline{r}}{256}) \times \Delta R^2 + 4 \times \Delta G^2 + (2 + \frac{255 - \overline{r}}{256}) \times \Delta B^2} \tag{6}$$

where

$$\overline{r} = \frac{C_{1,R} + C_{2,R}}{2} \tag{7}$$

$$\Delta R = C_{1,R} - C_{2,R} \tag{8}$$

$$\Delta G = C_{1,G} - C_{2,G} \tag{9}$$

$$\Delta B = C_{1,B} - C_{2,B} \tag{10}$$

Figure3 depicts segmentation results at 4 different levels.

## III. ROOFTOP DETECTION

The objective of this step is to generate a set of rooftop hypotheses based on the previously extracted straight lines and image segments.
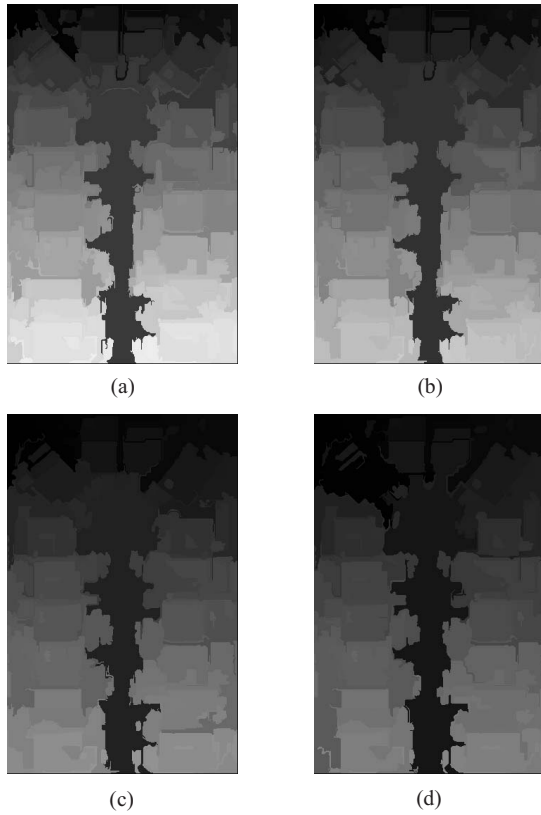
(a)                          (b)

(c)                          (d)

Fig. 3.   Segmentation results at 4 different sensitivity levels.

## A. Rooftop hypothesis generation

In this section, each segment is processed to create a set of possible rooftop hypotheses. The algorithm implemented to achieve this goal is as follows:

1. For each segment in the image segment dataset, apply morphological filtering by an opening followed by a closing with a structuring element of size $3 \times 3$.
2. Detect lines that describe the boundaries of each segment using the line detection algorithm described in II-A, II-B and II-C.
3. Identify segment's boundary pixels (Canny edge detector is employed in the implementation).
4. Find all the line segments in close vicinity with border points.
5. Combine the lines detected from steps 2 and 4 to generate one set of line segments associated with the current segment.
6. Generate a list of anti-parallel pairs of line segments. The direction of each line segment is computed based on the gradient direction of pixels.
7. For each pair of lines (denoted as West and East to follow the convention established in [14]), search for all approximately perpendicular line segments which could form North and South sides to the line pairs.
8. Generate a set of rooftop hypotheses from each combination of West, East, North, and South line segments
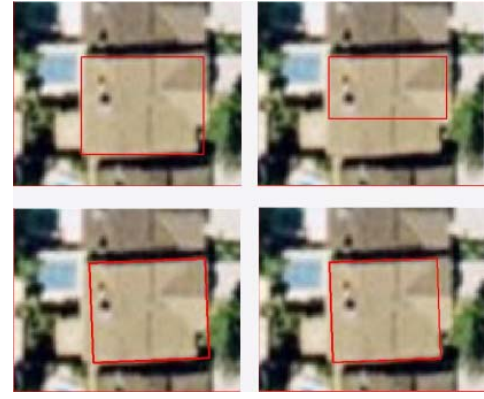


Fig. 4.   Four rooftop hypotheses initiated from one segment.

found.

Note that measures such as parallel, antiparallel, and perpendicular are determined in object space; however, with nadir imagery, it is assumed that angular measures taken in image space are approximately the same as those in object space. Figure 4 represents an example where all the hypotheses detected for one specific segment is over imposed on the original image.

## B. Hypothesis probability computation

In this section for each hypothesis a probability score is computed.

1. For each hypothesis create a tube around its sides.
2. Find all the lines that fully or partially fall within the tube. These lines are chosen from the set generated in Section III-A-5.
3. Remove lines with an orientation difference (from the side) higher than $15^o$.
4. Project each line on the closest side of the hypothesis.
5. For each side compute the normalized coverage due to the line projection in Step 4. Sum the four scores and normalize it.

Figure 5 represents a candidate hypothesis with all the lines that fall inside or intersect with the tube of 5 pixels around the rooftop definition.

## C. Local hypothesis refinement

The objective of this process is to inspect all hypotheses generated from one segment and choose the best hypothesis that could represents the area under inspection. In this process, from each set of overlapping hypotheses at most one hypothesis will be maintained. If the evidence for the best hypothesis is not strong enough, it will be eliminated. Following steps describe the refinement process:

1. Compute the normalized ratio of each hypothesis area covered by the initiating segment.
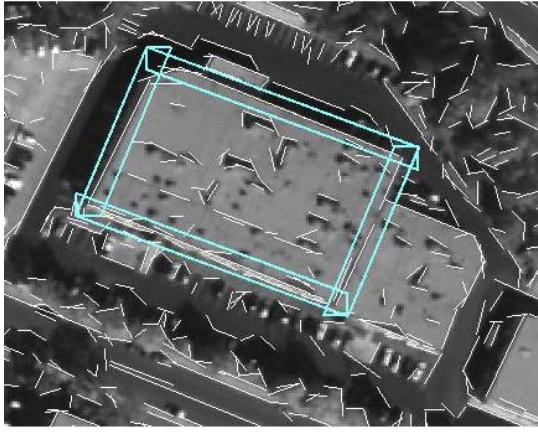2. Compute the percentage of the overlap between the segment and the area of the hypothesis.

Fig. 5. Line segments that fully or partially fall within the tubes (shown by blue lines) are projected onto the corresponding sides.

3. Combine each hypothesis probability with its coverage ratios (computed above). A simple mean is used in this work's implementation.
4. Choose the hypothesis with the highest score.
5. If the score is larger than a configurable threshold add this hypothesis to the hypothesis data set and otherwise remove it from the rest of process.

Steps in Sections III-A, III-B, and III-C are repeated for every segment in the segment data set.

Since each scene is segmented a number of times, there is a high possibility that for each physical building multiple hypotheses generated. Segmentation with different sensitivity parameter could cause partial or over segmentations. This issue could create overlapping hypotheses that partially or fully correspond to the same physical building but with different definition. Figure 6 shows all 72 hypotheses that are detected for the sample imagery.

Following steps are performed to eliminate false hypotheses or to choose the best representative hypothesis among two or more overlapping hypotheses.

### D. Removing pools and green areas in color images

The objective of this process is to eliminate hypotheses that have been generated over pools or yard areas. For this purpose the blue and green band images are processed to compute a Green-Blue ratio. This ratio is computed for each hypothesis by accumulating the hue value of the pixels inside each rooftop region. After normalizing this ratio by the area of the hypothesis, hypothesis with a ratio more than a predefined threshold is classified as an outlier and removed from the hypotheses candidate list. Figure 11.(a) shows the results before this step. Figure 11.(b) shows removal of a false hypothesis corresponding to the green area. The hypothesis in the mid-right section of image (a) is removed in image (b).

### E. Global hypothesis refinement

The objective of this process is to select the best hypothesis representing a building among a group of overlapping hypotheses.



Fig. 6. Hypotheses before filtering.
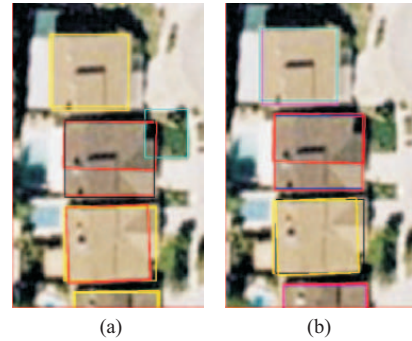


(a)                    (b)

Fig. 7. Hypotheses with high concentrations of Blue and Green are filtered.

1. The geometrical overlap is computed by measuring the intersection of convex polygons that define the rooftop hypotheses.
2. After all the geometric overlaps are computed, overlapping groups will be generated. If a hypothesis overlaps with image instances of two or more buildings, it will be moved to the group with the highest amount of overlap.
3. In each group the best representative will be chosen using the hypothesis probability measures computed in Sections III-B and III-C.

Figure 8 represents refined hypotheses shown Figure 8. The number of hypotheses is reduced from 72 to 17.

## IV. EXPERIMENTAL RESULTS

The proposed system is implemented in Matlab except for the segmentation routine that was written in C++ and is called from Matlab environment. The performance on an Intel

Fig. 8. Final results for scene 1, Los Angeles, CA, USA.



Fig. 9. Final results for scene 2, Los Angeles, CA, USA.

$\text{Core}^{TM}2$ Duo processor at $1.83$ GHz with 1 GB RAM for an image of $300 \times 200$ pixels is 363 seconds.

The performance of the system is assessed using many sample images from Google Maps, QuickBird satellite and aerial resources. Figures 8 to 13 illustrate the final results in each case. Note that partial buildings (at image border lines) are not processed and therefore have no correspondences in the presented results.

Table I describes a summary of the rooftop hypotheses from the above images.



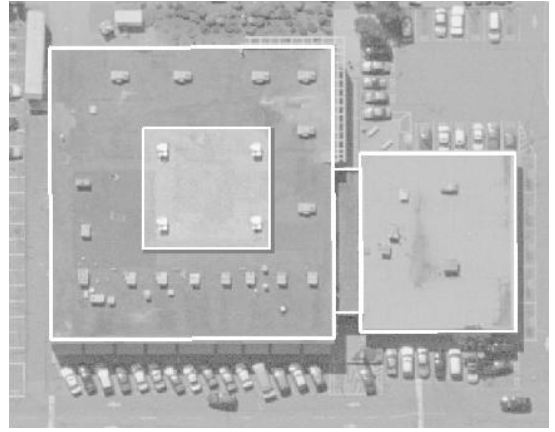Fig. 10. Final results for scene 3, Richmond, BC, Canada.



Fig. 11. Final results for scene 4, Richmond, BC, Canada.

In this table, BDP represents the Building Detection Percentage and QP stands for the Quality Percentage. These measurements were originally introduced by [18]. We have adjusted one of them.

$$BDP = \frac{100TP}{TP + FN} \tag{11}$$

$$QP = \frac{100TP}{TP + NTP + FP + FN} \tag{12}$$

## V. Conclusion

This paper described a system for detection of the building rooftops in satellite/aerial imageries. The system combines line definitions with segmented areas in the city images to detect buildings of various sizes. The system relies on blue band for color images (with an exception in Section III-D).

The plan for the future work is to improve the current system to incorporate refinement processes that inspect final results and adjust them accordingly. Generalizing the fitting algorithm for polygon shapes (other than quadrilateral) is currently under investigation. It is also planned to reduce the number of false negatives by means of a second process that runs on areas with high edge intensity and low hypothesis

Fig. 12.   Final results for scene 5, Ottawa, ON, Canada.



Fig. 13.   Final results for scene 6, Los Angeles, CA, USA.

correspondance. The long term goal for this project is to be incorporated in an automatic 3D building detection using monocular satellite imageries.

### REFERENCES

[1] Colour Metric. *http://www.compuphase.com/cmetric.htm*, 1996.
[2] A. Brunn and U. Weidner. Hierarchical bayesian nets for building extraction using dense digital surface models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 53(5):296–307, October 1998.
[3] J. Burns, A. Hanson, and E. Riseman. Extracting straight lines. *Trans. on Pattern Analysis and Machine Intelligence*, 8(4):425–455, July 1986.
[4] H. Cheng, Y. Chen, and X. Jiang. Thresholding using two-dimensional histogram and fuzzy entropy principle. *IEEE Trans. on Image Processing*, 9(4):732–735, April 2000.
[5] L. Davis. Hierarchical generalized hough transforms and line-segment based generalized hough transforms. *IEEE Trans. on Pattern Recognition*, 15(4):277–285, 1982.
[6] P. Felzenszwalb and D. Huttenlocher. Efficient graph-based image segmentation. *Int. Journal of Computer Vision*, 59(2):167–181, September 2004.
[7] A. Fischer, T. H. Kolbe, and F. Lang. Integration of 2d and 3d reasoning for building reconstruction using a generic hierarchical model, September 1997.
[8] C. Fraser, E. Baltsavias, and A. Gruen. Processing of ikonos imagery for submetre 3d positioning and building extraction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 56(3):177–194, April 2002.
[9] S. Heuel and T. H. Kolbe. Building reconstruction: The dilemma of generic versus specific models. *Knstliche Intelligenz, KI*, 15(3):57–62, 2001.
[10] R. Irvin and D. McKeown. Methods for exploiting the relationship between buildings and their shadows in aerial imagery. *Proc. SPIE*, 19(6):1564–1575, November 1989.
[11] C. Jaynes, E. Riseman, and A. Hanson. Recognition and reconstruction of buildings from multiple aerial images. *Computer Vision and Image Understanding*, 90(1):68–98, April 2003.
[12] S. Krishnamachari and R. Chellappa. Delineating buildings by grouping lines with mrfs. *IEEE Trans. Image Processing*, 5(1):164–168, January 1996.
[13] D. Lavigne, P. Saeedi, N. Goldstein, and H. Zwick. Automatic building detection and 3d shape recovery from single monocular electro-optic imagery. *SPIE Defense and Security Symposium*, 6567, June 2007.
[14] C. Lin and R. Nevatia. Building detection and description from a single intensity image. In *USC Computer Vision*, 1998.
[15] H. Mayer. Automatic object extraction from aerial imagery: A survey focusing on buildings. *Computer Vision and Image Understanding*, 74(2):138–149, May 1999.
[16] J. Peng and Y. C. Liu. Model and context-driven building extraction in dense urban aerial images. *Int. Journal of Remote Sensing*, 26:1289–1307(19), April 2005.
[17] S. Scholze, T. Moons, and L. Van Gool. A probabilistic approach to building roof reconstruction using semantic labelling. In *Proceedings of the 24th DAGM Symposium on Pattern Recognition*, page 257 ff., 2002.
[18] J. Shufelt. Performance evaluation and analysis of monocular building extraction from aerial imagery. 21(4):311–326, April 1999.
[19] G. Sohn and I. Dowman. Extraction of buildings from high-resolution satellite data. In *In Proceedings of the First Annual Conference of the Advanced School for Computing and Imaging*, pages 345–355, 2001.
[20] I. Suveg and Vosselman. 3d reconstruction of building models. In *In Proceedings of the XIXth ISPRS Congress, volume XXXIII-B3 of International Archives of Photogrammetry and Remote Sensing*, pages 538–545, 2000.
[21] U. Weidner and W. Forstner. Towards automatic building extraction from high-resolution digital elevation models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 50(4):38–49, August 1995.

TABLE I
SUMMARY OF THE DETECTION RESULTS.

| Scene Number | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| No. of Rows | 527 | 315 | 640 | 600 | 300 | 285 |
| No. of Cols | 329 | 293 | 320 | 400 | 500 | 188 |
| Total No of Buildings | 16 | 14 | 5 | 4 | 8 | 11 |
| Detected with 90% or more Accuracy (TP) | 13 | 11 | 4 | 4 | 7 | 8 |
| Detected with less than 90% Accuracy (NTP) | 3 | 2 | 1 | 0 | 0 | 1 |
| No. of FP | 0 | 1 | 0 | 0 | 0 | 0 |
| No. of FN | 0 | 1 | 0 | 0 | 1 | 2 |
| BDP (%) | 100 | 91 | 100 | 100 | 87 | 80 |
| QP (%) | 81 | 73 | 80 | 100 | 87 | 72 |