

# Automatic building detection and 3D shape recovery from single monocular electro-optic imagery

Daniel A. Lavigne<sup>\*a</sup>, Parvaneh Saeedi<sup>c</sup>, Andrew Dlugan<sup>b</sup>, Norman Goldstein<sup>b</sup>, Harold Zwick<sup>b</sup>

<sup>a</sup>Defence R&D Canada – Valcartier, 2459 Pie-XI Blvd, Quebec, QC, Canada G3J 1X5;

<sup>b</sup>MacDonald, Dettwiler & Associates, 13800 Commerce Parkway, Richmond, BC, Canada V6V 2J3;

<sup>c</sup>Simon Fraser University, School of Eng., 8888 University Dr., Burnaby, BC, Canada V5A 1S6

## ABSTRACT

The extraction of 3D building geometric information from high-resolution electro-optical imagery is becoming a key element in numerous geospatial applications. Indeed, producing 3D urban models is a requirement for a variety of applications such as spatial analysis of urban design, military simulation, and site monitoring of a particular geographic location. However, almost all operational approaches developed over the years for 3D building reconstruction are semi-automated ones, where a skilled human operator is involved in the 3D geometry modeling of building instances, which results in a time-consuming process. Furthermore, such approaches usually require stereo image pairs, image sequences, or laser scanning of a specific geographic location to extract the 3D models from the imagery. Finally, with current techniques, the 3D geometric modeling phase may be characterized by the extraction of 3D building models with a low accuracy level. This paper describes the Automatic Building Detection (ABD) system and embedded algorithms currently under development. The ABD system provides a framework for the automatic detection of buildings and the recovery of 3D geometric models from single monocular electro-optic imagery. The system is designed in order to cope with multi-sensor imaging of arbitrary viewpoint variations, clutter, and occlusion. Preliminary results on monocular airborne and spaceborne images are provided. Accuracy assessment of detected buildings and extracted 3D building models from single airborne and spaceborne monocular imagery of real scenes are also addressed. Embedded algorithms are evaluated for their robustness to deal with relatively dense and complicated urban environments.

**Keywords:** Building detection, 3D building modeling, 3D from 2D

## 1 INTRODUCTION

The extraction of 3D building geometric structures from high-resolution electro-optical imagery is nowadays one of the most complex and challenging tasks faced by the photogrammetry community, considering the scene complexity involved, crowded with objects of dissimilar natures, and cumulated with the diversity of scenario. It has been an active research topic, since producing 3D urban models is a requirement for varieties of applications such as urban planning, military simulation, creation of geographic information systems databases, and creation of urban city models. Furthermore, the emergence of commercially available high-resolution imagery has increased the need for such 3D geometric modeling capabilities, for application like topographic mapping<sup>1</sup>. Site modeling of dense urban areas can now be accomplished with the use of commercial spaceborne electro-optical sensors that provide imagery with spatial resolution below the half-meter level. However, while the reconstruction of 3D building models is a key element to generate the 3D geometric description of a scene, the extraction of 3D building models is traditionally achieved manually, which may result in extraction results with low accuracy level. Such manually based editing is time consuming, expensive, and requires well-trained operators. Moreover, actual techniques use digital terrain models, laser scanning, calibrated stereo pairs, or image sequences to help the operator in the extraction of the 3D building geometric shape<sup>2</sup>. However, in some practical cases, only a single electro-optic image of a given geographic location might be available<sup>3</sup>. Numerous techniques have been proposed in the past few years to automate the 3D building extraction process. Nevertheless, almost all 3D building reconstruction approaches are semi-automated ones<sup>4</sup>, limited to specific applications and restricted scenarios. Still, there is a growing interest in the development of automatic building extraction systems, aimed at the detection of 3D building models, with improved accuracy level, and able to ingest single monocular imagery acquired over complex urban scenes.

\*daniel.lavigne@drdc-rddc.gc.ca; phone 1 418 844-4000 x4157; fax 1 418 844-4511; www.drdc-rddc.gc.ca

## 2 EXTRACTION OF BUILDING STRUCTURES

The extraction of 3D building structures from remote sensing imagery is complex, mainly because it usually results in ambiguous solutions<sup>5</sup>. Indeed, the first difficulty experienced in developing automated building extraction techniques is the range of image variations in terms of type, scale, spectral range, sensor geometry, image quality, imaging conditions (e.g. lighting, weather), and the required level of detail. Using single monocular imagery increases substantially the level of difficulty to be considered, as buildings can be rather complex structures with many architectural details and a diversity of roof structures. A second difficulty arises with the automatic recognition of the semantic information embedded in the imagery, where the objects to be extracted in image-space can be located very close to each other. As a result, the buildings may be surrounded by numerous man-made and natural objects. This includes the occlusion of building parts and the geometrical resolution that may be limited. Suitable building extraction techniques must cope with the interpretation of such complex imagery, in order to recognize the location and extent of the buildings within these many image features. In spite of all these difficulties, some automated building extraction systems have been developed over the years, with limited performance in terms of robustness and accuracy.

### 2.1 Previous automatic building extraction systems

Previously developed automatic building extraction systems usually consider two main tasks within the building extraction process<sup>2</sup>: building detection and building reconstruction. These two tasks are mandatory, whether the extraction process is based on a specific model or if a priori information is used. For instance, building outlines and roof structures may be described with the use of lines and regions<sup>6</sup>, planar patches<sup>7</sup>, polyhedral shapes<sup>8</sup>, geometrical representation with rectangular models<sup>9</sup>, or using multiple images<sup>10,11</sup>. These proposed techniques are grouped as structural approach<sup>12</sup>, parametric models<sup>9</sup>, or perceptual organization<sup>13</sup>. While numerous semi-automated systems have been developed, a limited number of fully automated systems are found in the literature. Sohn and Dowman (2001) suggested an automatic building extraction technique using local Fourier analysis to analyze the dominant orientation angle in a building cluster in dense urban areas of IKONOS imagery<sup>14</sup>. Nonetheless, they assumed isolated buildings aligned parallel to a street without assessment of the accuracy of the modeling process. Fraser *et al* (2002) analyzed the potential of using high-resolution imagery for extracting building instances, by comparing extracted buildings from IKONOS imagery with the one extracted using airborne black and white images<sup>15</sup>. Using optimization and destruction approaches simultaneously, Otner (2002) introduced a point process technique to extract well-structured and symmetrical buildings<sup>16</sup>. Haverkamp (2003) used linking edges to extract buildings from IKONOS images<sup>17</sup>. Thomas *et al* (2003) evaluated three classification methods for the extraction of land cover information from high-resolution images<sup>18</sup>. Lee *et al* (2003) approximate the position and shape for candidate building objects in multispectral IKONOS imagery<sup>19</sup>. Shackelford and Davis (2003) used a pixel-based hierarchical classification to develop a preliminary estimate of potential buildings<sup>20</sup>. Benediktsson *et al* (2003) used mathematical morphological operations to extract structural information from the image<sup>21</sup>. Kim and Nevatia (2004) presented an approach for detecting and describing complex buildings, using expendable Bayesian networks to combine evidence from multiple images and digital elevation models<sup>11</sup>.

## 3 METHODOLOGY

The Automatic Building Detection (ABD) system aims at the automatic detection of buildings and the 3D reconstruction of their geometric structures from single monoscopic airborne or spaceborne panchromatic image view of a scene. Accordingly, the starting point of the process is strongly dependent of the type of input image data, scene complexity, sensor view angle, and on the state of preparedness of the input image as it is presented to the processing engine. Furthermore, insufficient ground sampling data and matching errors caused by poor image quality, occlusion, and shadows might lead to poor definition of buildings outlines.

In the development of the algorithms, the following assumptions are made:

1. In order to restrict the scope of the fundamental detection and 3D reconstruction problems, only panchromatic imagery are considered.
2. Both airborne and spaceborne imagery have full acquisition platform metadata supplied.
3. As no Digital Elevation Model (DEM) is provided, the surrounding terrain is assumed to be flat.

4. The building footprint is either rectangular or can be formed by simple composition of rectangles.
5. Within each rectangular building component, the walls of the buildings are vertical and have the same height.
6. The building roof types considered are flat(gable roofs are assumed to be flat at this time).
7. In order to have shadows as supporting evidence of 3D structures, the imagery is acquired on a clear day, with the sun neither at nadir nor at the horizon.

Under these assumptions, the proposed methodology is segmented into five major modules (Figure 1): 1) Acquisition geometry, 2) 2D image primitives, 3) Rooftop hypotheses generation, 4) Building hypotheses generation, and 5) 3D scene visualization. Each module is summarized in the following subsections.

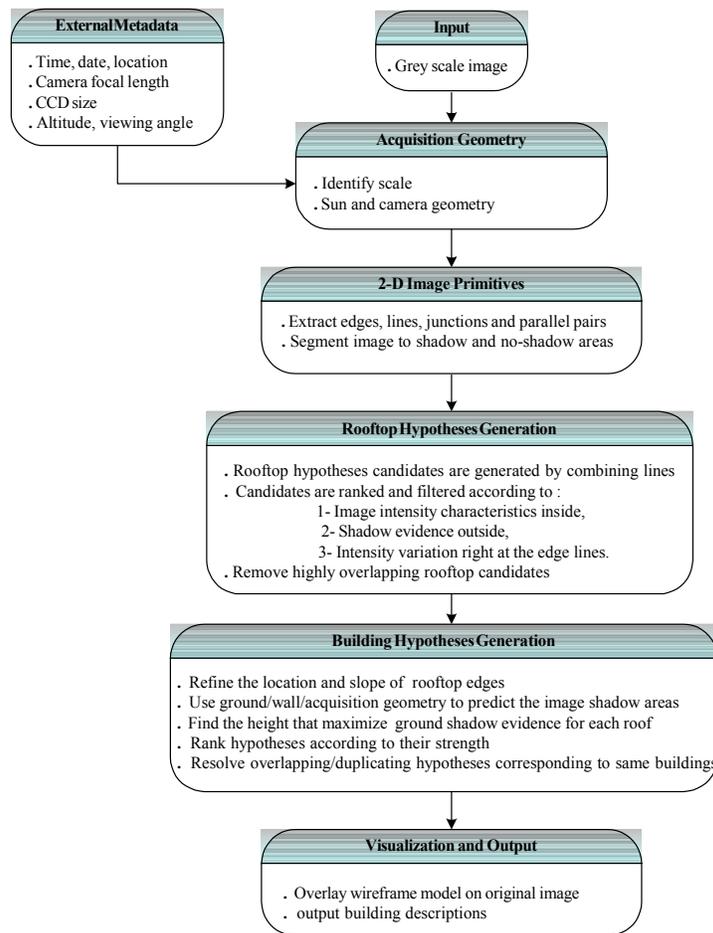


Figure 1. 3D building extraction process

### 3.1 Acquisition geometry

The acquisition geometry of diverse sensors and platforms is translated into data structures that are independent of the sensor and platform, and are well suited to efficient calculations during the image analysis. These data structures are vector fields that are essentially the jacobian of the ground to image and ground to shadow transformations. These transformations are slowly varying, so a sparse sampling is sufficiently accurate, and is more efficient at run-time than rational functions. The acquisition geometry is used to calculate the sun and sensor geometry, and identify the image scale. External metadata, such as acquisition time and sensor focal length, are provided to the module.

## 3.2 2D image primitives

Within this module, image primitives are extracted and processed from monocular panchromatic imagery. These primitives and processes, as listed below, are used in generating the initial set of building rooftop candidates:

- 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 linking process that links straight collinear line segments.
- A filtering process that removes lines that won't likely contribute to rooftop hypothesis generation.
- A segmentation of the image into shadow, and non-shadow regions. The non-shadow regions are also segmented into four distinctive regions according to their intensity values.

The following subsections describe the extraction of various image primitives.

### 3.2.1 Gradient field

The objective of this process is to generate the gradient field of the image that gives the gradient magnitude and direction for each image pixel. For this purpose, first a Gaussian lowpass filter is applied over the image to reduce pixel-level noise. Then the horizontal and vertical gradient at each pixel is computed by convolving vertical and horizontal gradient masks across the image. The gradient magnitude and direction are then computed from the horizontal and vertical gradients at each pixel point<sup>22</sup>.

### 3.2.2 Straight-line segments extraction

In this process, both gradient magnitude and gradient orientation of edge pixels are utilized to form line support regions and eventually straight-line segments<sup>22</sup>. The process is briefly summarized below:

1. Partition the 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 component 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 (i.e. original set and shifted set) of candidate lines.
6. For each line support region, compute the line represented by that region by performing least squares fit.

### 3.2.3 Linking process

The objective of this step is to link collinear line segments that are separated by very small gaps. The algorithm implemented in this step is as follows:

1. Sort the lines in the order they would be encountered if a horizontal sweep was performed across the image. This amounts to sorting lines by the leftmost endpoint coordinate.
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 of the following criteria must be satisfied for a pair of lines to be linked:
  - a. At least one of the lines must be longer than a supplied threshold value.
  - b. The lateral distance between the two lines must be lower than a threshold.

- c. The distance between the nearest endpoints of the two lines must be below a supplied threshold.
    - d. The slopes of the lines must be within an interval.
    - e. The degree of overlap (or under-lap) must be less than a supplied threshold.
  4. For each set of 2 or more lines that should be linked, replace them with a single line that spans the region previously covered by that set of lines. This replacement line can be generated in one of two ways:
    - a. Use the outside endpoints from the linked lines.
    - b. Merge the pixels from the line support regions into a larger line support region, and perform a least squares fit to compute the best-fit line. This latter method is the one used.
  5. Repeat steps 1 to 4 iteratively, modifying the linking thresholds on successive passes. While the “best” iterative strategy is configurable, one example is as follows:
    - a. Run the linking algorithm with *relatively* strict thresholds. The first pass will link up numerous shorter lines into longer ones.
    - b. Repeat the linking algorithm, without altering the thresholds. This second pass may further extend lines which were linked up in the first pass by extending to “outliers”.
    - c. Run the linking algorithm, but increase the length threshold while relaxing one or more of the other thresholds. This strategy aims to further extend the longer image lines (which tend to derive from building edges).

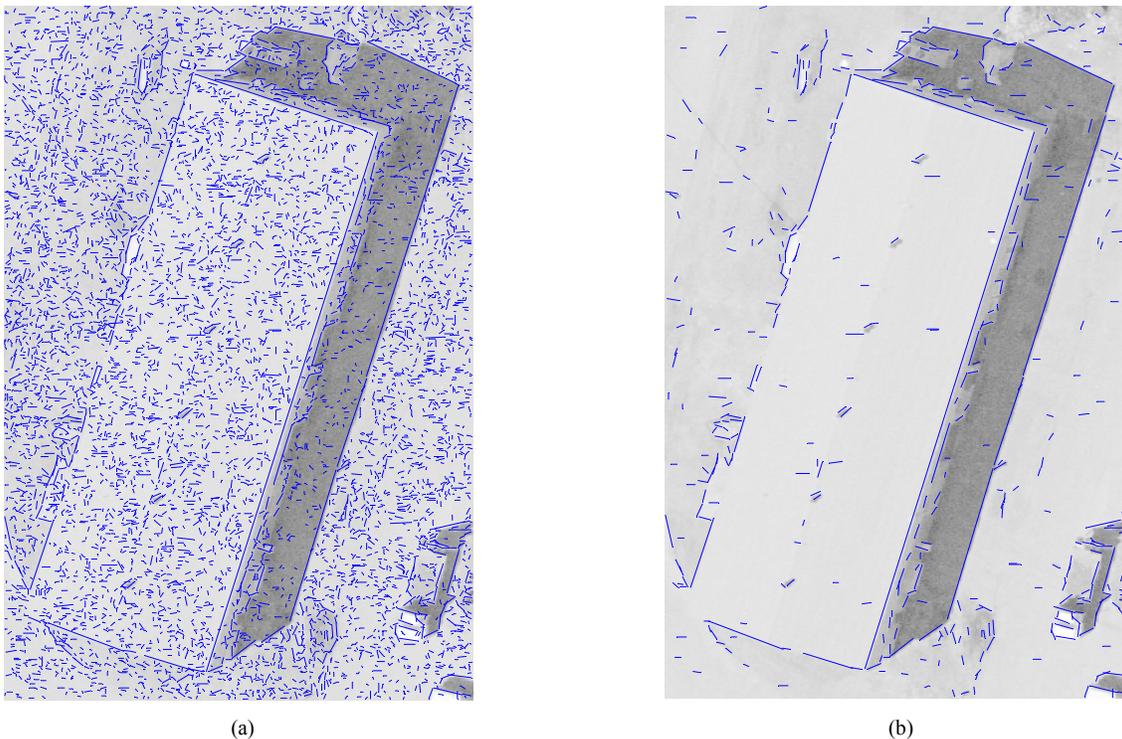


Figure 2. Line linking and filtering processes outputs of aerial imagery. (a) Line segments are extracted and then linked, according to some distance, orientation, slope, and overlap criteria. (b) Extracted lines are then filtered.

### 3.2.4 Line filtering

The objective of the line filtering process is to eliminate lines that are not likely to positively contribute to rooftop hypothesis generation. The motivation for doing this is that the execution time increases rapidly as the number of extracted lines increases. In theory, reducing the size of this line set allows subsequent analysis to be performed more rapidly without significantly impacting overall performance. Two filters are applied to the line set:

1. The first filter acts on line length. In general, “long” lines are more likely to contribute to hypothesis formation and subsequent steps. Thus, a filter was applied that removes all “short” lines, defined by a configurable parameter.
2. The second filter is slightly more complex. When the straight lines are produced from the line support regions, an average gradient across the line segment is calculated. Lines that separate regions of high contrast will have a large average gradient, while the ones that separate regions of low contrast will have a small average gradient. For each line, the average gradient is divided by the average intensity across the scene. This ratio is then compared to a configurable parameter. Lines with a ratio lower than this value are removed.

Figure 2a displays the results of line extraction after the linking process is applied, and figure 2b illustrates the results of line extraction after the linking process is applied.

### 3.2.5 Image segmentation

The image segmentation is performed to first identify areas corresponding to the shadow regions. The areas that do not correspond to shadow regions are further segmented into four groups. These later results are used when uniformity of each roof hypotheses is examined. Figure 3a shows the flow chart of the segmentation method applied to the input images, which is a modified version from the version that originally was described previously<sup>23</sup>. Figure 3b displays the resultant image after the segmentation

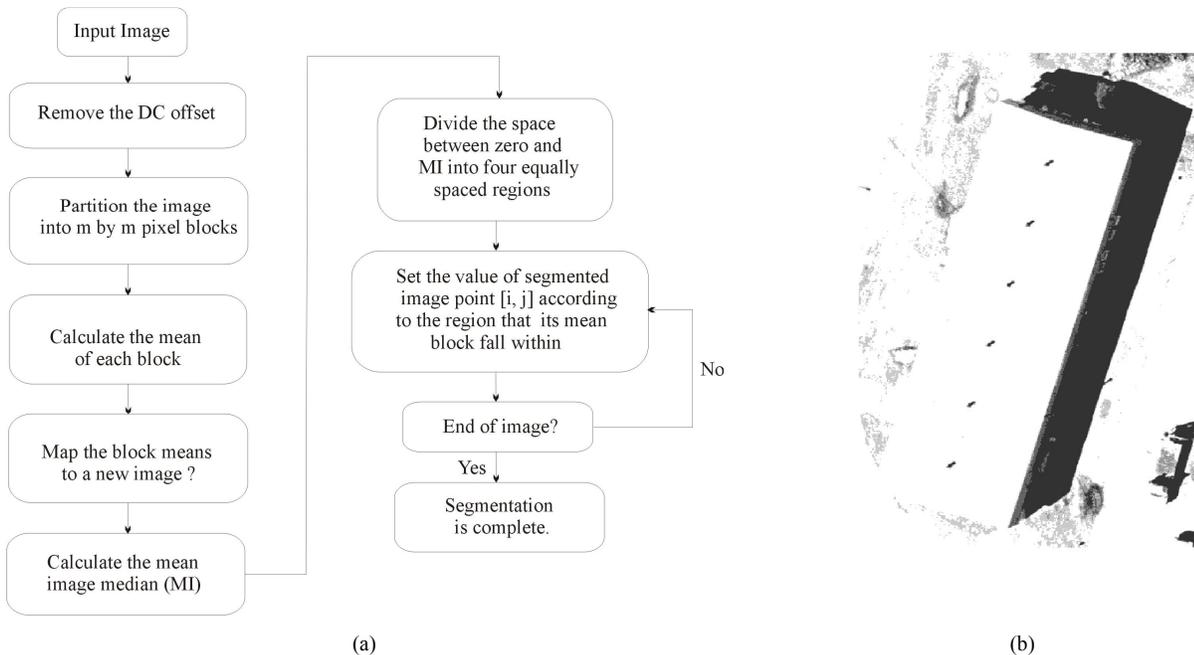


Figure 3. Segmented images into shadow and non-shadow regions. (a) Flow chart of the segmentation method. (b) Resultant segmented image into five regions.

### 3.3 Rooftop hypotheses generation

The objective of this module is to generate a set of rooftop hypotheses, based on the previously extracted straight-line segments. The initial set of roof hypothesis is extracted using straight lines set, then refined and assessed, based on their image intensity values. To achieve such roof hypotheses extraction, some algorithms have been developed for roof hypotheses generation and hypotheses refinement.

#### 3.3.1 Roof hypotheses generation

The algorithm implemented to achieve this is as follows:

1. Generate a list of anti-parallel pairs of line segments. When the line segments are extracted, they are associated with a certain direction based on the gradient direction of the pixels. Anti-parallel line segments have, by definition, gradient directions that differ by 180°.
2. For each pair of lines (denoted as ‘West’ and ‘East’, following a convention established previously<sup>22</sup>), search for all approximately perpendicular line segments which could form ‘North’ and ‘South’ sides to the building hypothesis.
3. Generate a rooftop hypothesis from each combination of West, East, North, and South line segments found.
4. For each hypothesis, determine all line segments that lie in the vicinity of the hypothesis. Use these line segments in the computation of a series of positive and negative hypothesis measures. Assign weights to each measure, and compute a score (weighted sum of the measures) for each hypothesis. These weights represent the likelihood that each rooftop hypothesis corresponds to a real rooftop.

Note that measures such as parallel, anti-parallel, and perpendicular must be determined in object-space. However, with almost nadir imagery, we assume that angular measures taken in image-space are approximately the same as those in object-space. Determination of the acquisition geometry of section 3.1 provides a method for converting between these two spaces.

#### 3.3.2 Roof hypotheses refinement

The initial set of roof hypotheses is further refined, by removing the potentially large number of “duplicate” hypotheses. Two hypotheses are assumed to be “duplicates” if:

1. The individual areas of the two hypotheses are comparably close.
2. The two hypotheses overlap significantly. To determine this, the area of the intersection of the two hypotheses is calculated, and then divided by the area of the smaller of the two hypotheses.

Choosing the true hypotheses among a set of possible duplicate is a very critical step and it would impact the accuracy of the ABD system to a large extend. For this purpose, the image content within each roof definition is incorporated in this process. The main assumption here is that true roofs seem to include more uniform image intensity distribution. Therefore, standard deviation of the intensity values bounded by each roof definition was computed next:

$$\delta_i = \sqrt{\frac{\sum_j \sum_k (I(j,k) - \bar{I})^2}{A}} \quad \text{Where } A \text{ is the area of image } I \quad (1)$$

All roof hypotheses with good geometric definition that pass the following condition will be passed to the next stage of the inspection:

$$\delta_i \leq (\min(\delta) + (\max(\delta) - \min(\delta))/2) \quad (2)$$

In order to identify the true rooftop definition, two more conditions are investigated:

- a. Transition from non-shadow to shadow for each roof edge.
- b. Evidence of shadow in the neighbourhood of roof definition.

These two conditions are applied to a shadow image that is generated earlier in Section 3.2.5. The first condition ranks each hypothesis according to the average number of pixels that change from non-shadow to shadow state when moving in the direction perpendicular to each roof edge ( $Trans_H$ ):

$$Trans_H = \frac{\sum_{l=1}^4 \sum_{p=1}^n |S_{l,p-1} - S_{l,p+1}|}{n} \quad (3)$$

Here  $S_{l,p-1}$  and  $S_{l,p+1}$  represent points at the two sides of line  $l$  (with length  $n$ ) of hypothesis  $H$  in the shadow segmented image  $S$ . Therefore, from a set of highly overlapped hypotheses, only the hypothesis with the strongest contrast across all four edges is selected and the rest of group candidates are retired from further assessments.

The second condition examines the evidence of shadow in the neighbourhood of each candidate hypothesis. For the true hypothesis, one expects to have shadow projections outside the two roof edges that in fact create such shadows. This condition removes false hypotheses that for instance are built upon lines that are generated at the intersection of shadow and ground. This condition is shown by:

$$\left( \sum_{i=-l-\Delta 1}^{l+\Delta 2} \sum_{j=-w-\Delta 3}^{w+\Delta 4} S_{i,j} - \sum_{i=-l}^l \sum_{j=-w}^w S_{i,j} \right) > 0 \quad (4)$$

In this equation, the first term represents the shadow area over the hypotheses extended (along both shadow edges). The values of  $\Delta 1$ ,  $\Delta 2$ ,  $\Delta 3$  and  $\Delta 4$  are defined based of the vector field. Figure 4 shows a typical example where three neighboring building rooftops, with a large interrelationship, are identified accurately.



Figure 4. Rooftops are identified for a scene with three buildings.

### 3.4 Building hypotheses generation

The objective of this module is to look for evidence of shadow, wall, and ground lines to provide three-dimensional support for each rooftop hypotheses. This allows a height prediction, and thus generates a 3D building hypothesis from the previously “flat” rooftop hypotheses. Correct height estimation requires accurate roof localization. For this reason, an extra step is taken here to refine the location and slope of rooftop definitions.

### 3.4.1 Roof location and slope refinement

Positional refinement of the roofline hypothesis requires analyzing image intensities around each candidate line/edge. By definition these areas, if identified correctly, must contain abrupt changes or discontinuities in the image properties such as intensity, color or texture. This discontinuity can be noticed when moving on the perpendicular direction at each edge point. However, this is under the assumption that an accurate edge definition (slope and position) exists. In this process, we first correct the slope of each rooftop line and then refine its location. With the assumption that the true edge locates in a close vicinity of the current definition, each edge is re-sampled (configurable in width and method) around its current definition first. Next, the re-sampled image is Gaussian filtered. The Laplacian operator is then applied to the resultant image region. By definition, the edge is located at the point where the Laplacian sign changes. A fit is then computed that minimized the edge location error along the line direction: this fit represent the corrected roof edge definition. After the slope of each roofline is corrected, it is re-sampled once again. A directional gradient (perpendicular to the edge direction) is then applied to the image region. Using the one-dimensional profile of the resulting gradient, the location of the edge is then computed, first by looking at the maximum location. It is then refined by sub-pixel interpolation using a 2<sup>nd</sup> order polynomial fit through the following equation:

$$\Delta x = (y(x+1) - y(x-1)) / (2 \times (2 \times y(x) - y(x+1) - y(x-1))) \quad (5)$$

The edge refinement may be chosen to apply on all four edges, or just simply to the edges that produce the shadow footprints. In our experiments, there was not noticeable changes (within a pixel) when refining edge locations using “linear” and “nearest” interpolation methods. This could be due to the fact that the one-dimensional profile was smoothed by the mean filter. Also, the sub-pixel interpolator of equation (5) still provides accuracy within a pixel, which seems to be sufficient for the examined cases.

### 3.4.2 Height estimation

Estimating height of buildings could be performed by matching the projected shadows of a set of candidate heights, against the extracted lines around each projection neighborhood. This approach performs robustly in the case of buildings with clear shadows and no occlusions. In cases where shadows are partial due to objects occluding the shadow lines or adjacent buildings, where the shadow lines are not clearly drawn, it could fail (as shown in Fig. 5a).

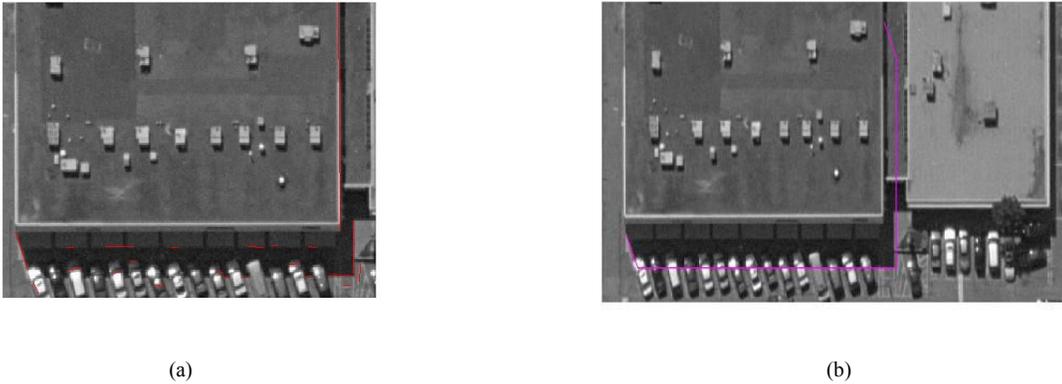


Figure 5. Height estimation process. (a) Partial shadow lines result from occluding objects. (b) Projected shadows of a set of candidate heights are matched against the extracted building lines.

For the above reasons, a complementary algorithm is implemented that takes advantage of the image intensity values around the shadow line. Generally moving to or from a shadow region produces a large intensity change. Therefore, for every possible height, projected shadow lines are estimated first. A building could have up to 4 shadow lines. The mean normalized intensity variation in the direction perpendicular to each shadow line is computed next. The pixel variation is weighted according to the length of each edge. Finally, scores for all edges are combined to compute an overall score: the height with the highest score is chosen as the height of the building.

Since this method combines shadow evidences from all roof edges, it performs more robustly under occlusion. For instance, if parts of shadow lines are lost due to objects/adjacent building (as shown in Fig. 5b), there is still a high possibility that the remaining parts would sufficiently characterize the correct height of the building.

### 3.4.3 Building hypotheses selection

Separation of the true hypotheses from faulty ones is a key element in the reliability of the ABD system. For this reason, we combine the 3D score of each building with those of the roof definition to generate support ratio for each building hypotheses.

The 3D score of the building hypotheses are estimated using the projection of all shadow lines on the image. The image is re-sampled at the location of these edges and the mean relative intensity variation for each individual edge is computed using:

$$\Delta I_{i,j} = [I_{i,j-1} \quad I_{i,j} \quad I_{i,j+1}] \cdot [-1 \quad 0 \quad 1] \quad (6)$$

Here  $I_{i,j}$  represents the  $i^{\text{th}}$  pixel on the  $j^{\text{th}}$  shadow line.

$$\mu_n = \frac{\sum_{i=1}^m \Delta I_{i,j}}{m} \quad \text{where } i, j \in \text{edge } n \text{ of length } m \quad (7)$$

This value is weighted according to the edge length. Values corresponding to all edges of each building rooftop are added together to generate a shadow edge intensity score:

$$HS = \frac{\sum_{n=1}^4 \mu_n \times L(n)}{\sum_{n=1}^4 L(n)} \quad (8)$$

The overall support ratio for each hypothesis is then computed using the following equation:

$$\text{Score} = RS \times HS - k \times (RS + HS)^2 \quad (9)$$

In this equation,  $RS$  is the rooftop score and  $k$  is an empirical constant<sup>24</sup>, which is set to 0.04.

### 3.4.4 Resolving overlapping hypotheses

The objective of this step is to resolve any remaining inconsistencies in the building hypothesis set by handling pairs of hypotheses, which overlap significantly. Following sequence is performed for each pair of building hypotheses:

- a. Determine the overlap ratio between the two hypotheses. As before, this is taken as the area of the intersection between the two rooftops divided by the area of the smaller rooftop.
- b. Hypotheses are deemed to be “overlapping” if the overlap ratio exceeds a predefined threshold. Ideally, this ratio threshold should be set to zero since real 3D buildings should not overlap at all. However, to account for variations in the building extraction process, a small tolerance is allowed.
- c. Whenever two building hypotheses are deemed to overlap, the hypothesis with the lower overall building score is eliminated.

## 4 EXPERIMENTAL RESULTS

The ABD system was tested with a set of 12 aerial and satellite images: two experimental results from the data set are presented in this section. Figure 6a shows the results for a scene including 3 neighboring buildings, together with a summary of the building hypothesis scores, including the estimated heights for each building (Fig. 6b). The predicted rooftop lines are drawn in solid red, the predicted wall/ground lines (if applicable) are shown as dashed red lines, and the

predicted shadow lines are shown as dashed blue lines. Figure 7 illustrates another experimental results: 16 building hypotheses generated from a dense suburban area.



(a)

“True” Buildings?	ID	Hypothesis Score	Height
Yes	943	0.0090	5.5
Yes	8190	0.0081	3.5
Yes	8520	0.0333	3

(b)

Figure 6. 3D building models extracted from an airborne image. (a) Extracted buildings are displayed with wireframes. (b) Hypothesis score and estimated building heights.



Figure 7. 3D building models extracted from a dense suburban area.

## 5 CONCLUSIONS AND FUTURE WORK

This paper described a system that actually is in development for the automatic building detection and extraction of 3D models from single monocular high-resolution electro-optical imagery. First, the ABD system automatically extracts 2D image primitives such as edges, lines, junctions, and parallel pairs, while providing a segmentation of the image in shadow and no-shadow areas. Rooftop hypothesis candidates are then generated by combining lines previously extracted, and then ranked and filtered according to a set of rules. Next, refinement of the location and slope of rooftop edges is executed. Building heights that maximize shadow evidences are estimated subsequently. Finally, building hypotheses are ranked, and 3D building models are generated. Currently developments are under progress to identify buildings on top of other ones. Re-weighting evidences based on building interactions (i.e. occlusion) are also in progress. Future work includes that if potential evidence is impossible because of occlusion, that aspect of the building's weights shall not be penalized so severely. Finally, shadow analysis could be recalculated based on the building interactions, providing an iterative approach to the analysis, proceeding until all the evidence is consistent with the 3D building model.

## REFERENCES

1. T. Toutin, R. Chénier, and P. Cheng, “3D models for high resolution images: examples with QuickBird, IKONOS and EROS”, *Symposium on Geospatial Theory, Processing and Applications*, Ottawa, Canada, 2002.

2. S. Noronha and R. Nevatia, "Detection and description of buildings from multiple aerial images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 5, pp. 501-518, 2001.
3. E.P. Baltsavias, "Object extraction and revision by image analysis using existing geospatial data and knowledge: current status and steps towards operational systems", *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 58, pp. 129-151, 2004.
4. A. Grün, "Semi-automated approaches to site recording and modelling", *International Archives of Photogrammetry and Remote Sensing*, Vol. 33, (5/1), pp. 309-318, 2000.
5. A. Brunn and U. Weidner, "Extracting buildings from digital surface models", *IAPRS*, Vol. 32, Part 3-4W2, 3D Reconstruction and Modeling of Topographic Objects, Stuttgart, September 17-19, 1997.
6. A. Fisher, T.H. Kolbe, and F. Lang, "Integration of 2D and 3D reasoning for building reconstruction using a generic hierarchical model", In *Proceedings of the Workshop on Semantic Modeling for the Acquisition of Topographic Information from Images and Maps*, SMATI, pp. 159-180, 1997.
7. B. Ameri and D. Fritsch, "Automatic 3D building reconstruction using plane-roof structures", *ASPRS 2000 Annual Conference*, Washington, DC, May 22-26, 2000.
8. S. Scholze, T. Moons, and L. Van Gool, "A probabilistic approach to roof extraction and reconstruction", In *Proceedings of the ISPRS Commission III Symposium*, Graz, Austria, 2002.
9. U. Weidner and W. Forstner, "Towards automatic building extraction from high-resolution digital elevation models", *International Society of Photogrammetry and Remote Sensing*, Vol. 50, No. 4, pp. 38-49, 1995.
10. C. Ballard and A. Zisserman, "A plane-sweep strategy for the 3D reconstruction of buildings from multiple images", *International Archives of Photogrammetry and Remote Sensing*, Vol. XXXIII, Part B2, pp. 56-62, 2000.
11. Z.W. Kim and R. Nevatia, "Automatic description of complex buildings from multiple images", *Computer Vision and Image Understanding*, Vol. 96, pp. 60-95, 2004.
12. F. Fuchs and H. Le Men, "Efficient subgraph isomorphism with a-priori knowledge: application to building reconstruction for cartography", *Lecture Notes in Computer Science*, Springer, 2000.
13. R. Nevatia and K. Price, "Automatic and interactive modeling of buildings in urban environments from aerial images", *IEEE International Conference on Image Processing*, New York, pp. 525-528, 2002.
14. G. Sohn and I. Dowman, "Extraction of buildings from high-resolution satellite data", In *Automated Extraction of Man-Made Objects from Aerial and Space Images (III)*, Swets & Zeitlinger B.V., Lisse, The Netherlands, pp. 345-355, 2001.
15. C.S. Fraser, E. Baltsavias, and A. Gruen, "Processing of IKONOS imagery for submetre 3D positioning and building extraction", *International Journal of Photogrammetry and Remote Sensing data*, Vol.56, pp.177-194, 2002.
16. M. Oter, X. Decombes and J. Zerubia "Building extraction from digital elevation model", *Technical Report INRIA No. 4517*, INRIA, France, 2002.
17. D. Haverkamp, "Automatic building extraction from IKONOS imagery", *ASPRS 2004 Annual Conference*, Denver, Colorado, May 23-28, 2004.
18. N. Thomas, C. Hendrix, and R.G. Cogalton, "A comparison of urban mapping methods using high-resolution digital imagery", *Photogrammetric Engineering and Remote Sensing*, Vol. 69, No. 9, pp. 963-972, 2003.
19. D.S. Lee, J. Shan, and J.S. Bethel, "Class-guided building extraction from IKONOS imagery", *Photogrammetric Engineering and Remote Sensing*, Vol. 69, No. 2, pp. 143-150, 2003.
20. A.K. Shackelford, C.H. Davis, "A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 41, No. 10, pp. 2354-2363, 2003.
21. J.A. Benediktsson, M. Pesaresi, and K. Arnason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 41, No. 9, pp. 1940-1949, 2003.
22. C. Lin and R. Nevatia, "Building detection and description from a single intensity image", *Computer Vision and Image Understanding*, Vol. 72, No. 2, pp. 101-121(21), 1998.
23. J.M. Scanlan, D.M. Chabries, R. Christiansen, "A shadow detection and removal algorithm for 2-D images", *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Albuquerque, New Mexico, pp. 2057-2060, 1990.
24. C. Harris , M. Stevens, "A combined corner and edge detector", *Proceedings of The Fourth Alvey Vision Conference*, Manchester, pp 147-151. 1988.