Fusion of Lidar and Imagery for Reliable Building Extraction

Dong Hyuk Lee, Kyoung Mu Lee, and Sang Uk Lee

Abstract

We propose a new building detection and description algorithm for lidar data and photogrammetric imagery using directional histograms, splitting and merging segments, and line segments matching. Our algorithm consists of three steps. In the first step, we extract initial building regions from lidar data. Here, we apply a modified local maxima technique coupled with directional histograms and the entropies of these histograms. In the second step, given the color segmentation results from the photogrammetric imagery, we extract coarse building boundaries based on the lidar results with region segmentation and merging from aerial imagery. In the third step, we extract precise building boundaries based on the coarse building boundaries using line segments matching and perceptual grouping. Experimental results on multisensor data demonstrate that the proposed algorithm produces accurate and reliable results.

Introduction

In recent years, the increasing need for accurate three-dimensional (3D) data of urban areas, and their continuous update has led to research efforts that aim to develop automatic or at least semiautomatic tools for the acquisition of such data. To satisfy the new demands, more automated methods that produce accurate geo-information are required to keep costs within reasonable bounds. With its high pulse frequencies, light detection and ranging (lidar) is a very valuable data source for the production of geo-information. For this reason, lidar plays an important role in both the automation of building detection and the creation of 3D topographical databases.

Early researchers generally used three kinds of methods to detect and reconstruct buildings from lidar data and photogrammetry. The first technique is to use only lidar data, because the photogrammetry of the region that corresponds to the region of the lidar data maybe impossible to obtain. There have been several attempts to detect building regions from lidar data. The task has been solved by classifying the lidar points according to whether they belong to bare-earth, buildings, or other object classes. Morphological opening filters are used to determine a Digital Terrain Model (DTM) that can be extracted by subtracting object points from the Digital Surface Model (DSM). By applying height thresholds to the normalized DSM, an initial building region is obtained. This initial classification must be improved to delete vegetation regions. In Brunn and Weidner (1997), this was accomplished by a framework for combining various cues in a Bayesian network. In Rottensteiner and Briese (2002), an algorithm for building detection that relied on DTM generation by hierarchic robust linear prediction (Briese *et al.*, 2002) was presented. The DTM and DSM grids in Rottensteiner and Briese (2002) were used for further classification.

The second type of method is to use 2D or 3D information from photogrammetric imagery. Early researchers tried to extract feature-ground separation, but it is hard to separate building boundaries from other distracting lines, such as road boundaries, by using only perceptual information such as parallel lines and right-angled corners. Therefore, other information is used for building extraction. The most notable is the depth information from stereo by multiple images (Sun *et al.*, 2005). Kim and Nevatia (2004) proposed a method that automatically constructs the description of complex buildings from multiple images. The main difficulty in utilizing stereo information is that although range data can be generated from stereo analysis, its quality is not good enough to generate building hypotheses directly, since the roofs of many buildings lack sufficient texture for stereo processing. In addition, nearby trees of similar height also make the use of such range data difficult.

The third type of methods uses both the lidar data and photogrammetric imagery. Since lidar and photogrammetric imagery each has unique advantages and disadvantages for reconstructing building surfaces, advantages of one method can compensate for disadvantages of the other method making it natural to combine the two methods. More specifically, intensity and height information in lidar data can be used with texture and boundary information in photogrammetric imagery to improve accuracy. Shenk and Csatho (2002) proposed feature-based fusion of lidar data and digital aerial images to obtain a better surface description than could be achieved by using only one of these data sources. Habib *et al.* (2005) proposed a method based on the registration of photogrammetric imagery and lidar data using linear features. Sohn and Dowman (2003) focus on an exploitation of synergy of Ikonos imagery combined with a lidar DEM. Specifically, individual buildings are localized with rectangle polygon by a hierarchical segmentation of lidar DEM and Ikonos multi-spectral information. However, this method has building extraction errors, such as intrusion/extrusion of building shape. The Rottensteiner *et al.* (2004) method consists of building detection step, roof plane detection step, and the determination of roof boundaries step. Building detection is based on the Dempster-Shafer theory for data fusion. In roof plane detection, the results of

Department of Electrical Engineering ASRI, Seoul National University, Korea (kyoungmu@snu.ac.kr).

Photogrammetric Engineering & Remote Sensing
Vol. 74, No. 2, February 2008, pp. 215–225.

0099-1112/08/7402–0000/$3.00/0

a segmentation of laser scanner data are improved using the digital images. The geometric quality of roof plane boundaries can be improved at step edges by matching the object edges of the polyhedral models with image edges. However, it is possible that true straight roof plane boundaries are changed as wrongly curved boundaries by matching and merging outlier edge segments. In this paper, we present a new approach to detect and describe complex buildings by using lidar data and aerial images.

System Overview

Lidar and photogrammetric imagery each have particular advantages and disadvantages in horizontal and vertical position accuracy. Compared with photogrammetric imagery, lidar provides accurate height information but inaccurate boundary lines. Unfortunately, some regions in lidar data have null values due to self-occlusion of a building. Photogrammetric imagery provides extensive 2D information such as high-resolution texture and color information. Although 3D height information can be estimated from one or several images by several methods such as stereo, shape from shading compared with lidar, this information is relatively inaccurate.

To extract accurate building region extraction results, we combine the information from lidar and photogrammetric imagery by first assuming the building regions extracted from lidar as the initial regions, and then improving the inaccurate boundaries by utilizing the photogrammetric imagery to obtain the precise building boundaries. More specifically, we first extract the initial building boundaries from the height information provided by lidar data. Next, we enhance the initial boundaries using color information provided by the photogrammetric imagery. Here, we apply color segmentation based on the assumption that building roofs are planar, and therefore, the height of the building and its color depicted in the photogrammetric imagery is homogeneous. Finally, we apply edge matching and closed loop construction as a post-process, since most man-made objects are best described by edges. Figure 1 shows the block diagram of the system.

Extracting Building Boundaries from Lidar

The extraction of building segments from lidar data is performed through two classification steps. First, the lidar points are classified into bare-earth points and building candidate points. Next, the vegetation points are identified and eliminated from the building candidate points. Both classification steps are performed on segmented lidar point clouds.

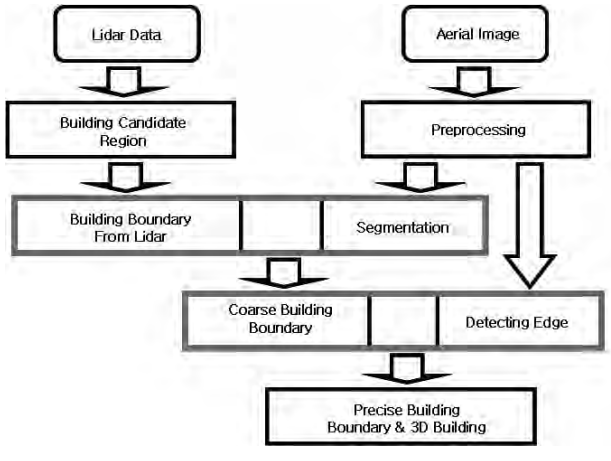


Figure 1. Block diagram of the system.

Classification of Bare-earth Points and Building Candidate Points

Classification of Each Point in Local Patch

We first subdivide the scanned regions into local patches where the patch size is predefined. A suitable patch size is determined based on these criteria plus a maximum building size constraint. Classification of candidate building points is done by local maxima filter that measures the difference of each point in the patch with the minimum height point of that patch.

Conversion of Candidate Building Regions into Grid Form

The data points in extracted building regions are converted to a grid form for easier/faster data processing. Commonly used interpolation techniques include bilinear interpolation and nearest-neighbor interpolation. In order to decide which interpolation technique to apply, we must consider the common attributes of artificial objects. Generally, bilinear interpolation does not preserve the discontinuities at the building boundaries and wrongfully smoothes the height of these points. On the other hand, nearest neighbor interpolation would cause errors in points where height is smoothly varying, as in the case of a sloped roof. However, there are not many un-scanned points in sloped roof regions. Difficulties in scanning the height are mainly caused by abrupt discontinuities of height concentrated at building boundaries. Therefore, nearest neighbor interpolation is more suitable than bilinear interpolation for this problem.

Deleting Vegetation Regions

Next, candidate building regions extracted in the previous process are further classified as building or vegetation. Traditionally, this further classification has been based on one or more of the following criteria for each region: spectral signature, area, circularity, and surface roughness or surface gradients. However, due to their simplicity, each attribute mentioned above has limitations. First, the use of spectral signatures is restricted by time and weather. Moreover, common shape parameters, such as roundness and area, have proved to be insufficient in this application to distinguish between vegetation and buildings. Finally, since it is possible for a building to have different heights in its roof, due to various installations such as water tank, chimney, antenna, dome or stair shape structures, simply using the gradient may cause problems such as removing partial regions of the roof that have different height. In this paper, we propose a new algorithm for deleting vegetation as follows. First, as shown in Figure 2, we integrate all candidate building points into groups depending on whether they belong to the same building using a region growing segmentation technique.

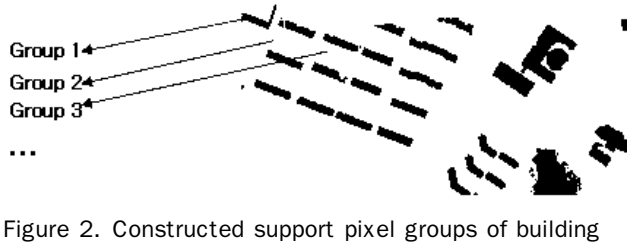


Figure 2. Constructed support pixel groups of building candidates after the region growing process.

Through this process, we are able to estimate the boundary of each building candidate region, and thereby, we are able to extract the characteristics of each region. From the extracted boundaries, we distinguish vegetation from buildings using second-order gradients and its directional histograms for each pixel. The following paragraphs provide details on each process.

After extracting building candidate points, binary values 0 and 255 are assigned to building candidate points and bare-earth points, respectively. This can be viewed as constructing a binary image from the preliminary classification results obtained from the previous process. Region growing is applied to this binary image to cluster building candidate points that are likely to belong to the same building. Specifically, region growing is applied to building candidate points, i.e., pixels with 0 values only, where the seed pixels for the process are randomly selected from 0 value pixels which have not been processed. Region growing is performed until all 0 value pixels have been processed, and all independent regions after processing are assumed to be separate buildings with each building region having a corresponding support pixel group.

Next, we distinguish regions of vegetation by using the second-order gradient and its directional histogram for each pixel. Building roofs that have non-zero gradient values, such as stair shaped roofs, slanted roofs, or roofs with chimneys have relatively few height values and correspondingly, few gradient orientations. However, vegetation regions have many different height values, ranging from very small to very large, and orientations of the gradients are generally unbiased and have many directions. Based on this observation, the first-order gradient can be used to discriminate between building regions and vegetation regions. Here, we use the second-order gradient due to dome structured buildings, since the second-order gradient values are fixed for domes. The second-order gradient magnitudes and orientations are calculated as follows:

$$M = \sqrt{x^{n2} + y^{n2}} \tag{1}$$

$$\theta(x, y) = tan^{-1} \frac{y^n}{x^n}, \tag{2}$$

where M is magnitude, θ is orientation, x'' is the second-order derivative of the x value and y'' is the second-order derivative of the y value.

The directional characteristic of a group is formed from the second-order gradient orientations at all pixels within the group, as shown in Figure 3b. These pixels are then accumulated into directional histograms summarizing the contents over the grouped region, as shown in Figure 3. Here, the directional characteristic is represented by a directional histogram representing the cumulative orientation of all pixels in the current group, where the length of each arrow in the histogram denotes the sum of the magnitude of all pixels with orientation corresponding to the direction of that particular bin. The directional histogram has 36 bins covering the 360-degree range of rotations, and peaks in the directional histogram correspond to dominant orientations of a group. In a building region the distributed orientations of second-order gradients have low complexity, while in a vegetation region, the complexity is high. Since it denotes the complexity of each group, the concept of entropy can be applied as the characteristic of complexity:

$$entropy(x) = -\sum_{i=1}^{i=36} p(x_i)\, log\, p(x_i). \tag{3}$$

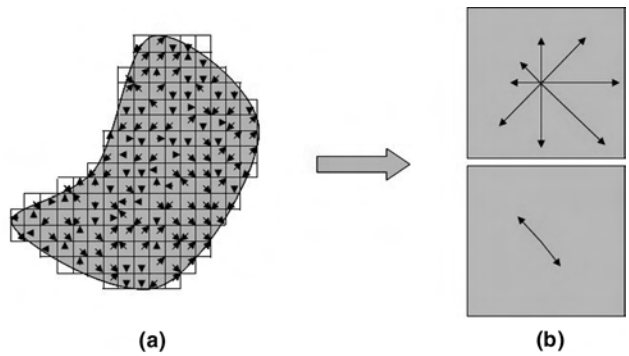


Figure 3. (a) The second order gradient magnitude and orientation at each pixel in the building region, and (b) a typical histogram of a vegetation region in the top, and a typical histogram for a building region in the bottom.

Using Equation 3, the entropy value is estimated for each group. In this equation, x_i denotes the value of each bin covering a 10-degree range in the histogram. $P(x_i)$ is a probability of bin i: the ratio of the value of the i^{th} bin to the total sum of all bins of the histogram.

To reiterate the important observations made, there are only a small number of orientations in a building roof regardless of whether the roof is flat or sloping as shown in the bottom-right histogram in Figure 3b; thereby, the histogram of building regions are restricted to have non-zero values for only a few bins. In this case, the estimated entropy will be low (in this case, the entropy is about 1). However, as shown in the top-right histogram of Figure 3, many directions are shown in a vegetation region, because the orientation distribution in vegetation is irregular and complex (in this case, the entropy is about 3). Therefore, a high entropy will be estimated, and consequently enable us to utilize this measure to distinguish vegetation regions from building regions. Note that if the building roof shape is a perfect planar surface, all of the second-order gradients are 0, and therefore Equation 2 is not defined. For this case, we simply set the entropy as 0, since the complexity of this case is the lowest. However, we didn't encounter exactly planar surface cases in our experiments, since we consider a whole building region to estimate the entropy, and generally, building roofs have some height variation.

Extracting Coarse Building Boundaries

Any automatic or manual point-matching method outputs some error in the process of registering data captured by photogrammetric and lidar systems to a common reference frame. Moreover, hills and trees around buildings or height variations in some buildings can disturb accurate extraction of building boundaries from lidar data. Therefore, before extracting precise building region, coarse building regions are extracted from the photogrammetric imagery using an over-segmentation and segment merging technique together with the results obtained based on the lidar data from the previous process.

Registration

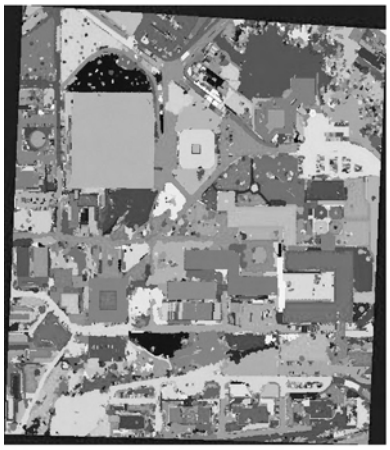
In this paper, we use high-resolution aerial images as photogrammetric imagery and assume that there are no internal/external orientation parameters available for the aerial image and lidar data. Therefore, we estimate homography for the

registration of lidar and aerial image (Hartely and Zisserman, 2003). We begin with a simple linear algorithm for determining H (homography) with a manually given a set of 2D aerial image to 2D lidar image point correspondences, $\mathbf{x}_i \leftrightarrow \mathbf{x}_i'$. The transformation is given by the equation $\mathbf{x}_i' = H\mathbf{x}_i$. By denoting $\mathbf{x}_i' = (x'_i, y'_i, w'_i)^T$, the homography equation involving homogeneous vectors may be written in the form:

$$\begin{bmatrix} 0^T & -w_i'\mathbf{x}_i^T & y_i'\mathbf{x}_i^T \\ w_i'\mathbf{x}_i^T & 0^T & -x_i'\mathbf{x}_i^T \\ -y_i'\mathbf{x}_i^T & x_i'\mathbf{x}_i^T & 0^T \end{bmatrix} \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} = 0. \quad (4)$$

These equations may be rewritten as $A_i h = 0$, where A_i is a 3×9 matrix, and h is a nine-dimensional vector made up of the entries of the matrix H. To solve the homography H, we provide a set of point correspondences manually. Figure 4a shows an image warped using a homography.

(a)



(b)

Figure 4. Preprocessing for coarse building boundaries: (a) warped image by homography, and (b) segmentation result by the mean-shift method. A color version of this figure is available at the ASPRS website: **www.asprs.org**.

Color Segmentation

Most conventional color segmentation algorithms are based on the information in the color space. However, clustering in color space only usually does not provide satisfactory performance, since it lacks information about the spatial configuration. To resolve this limitation, in addition to color, the spatial coordinates of pixels are often incorporated into their feature space representation. Here, we apply the mean-shift color segmentation method (Comaniciu and Meer, 2002) to obtain over-split segmentation results. As a result of this segmentation, one building region is divided into one or more segments. Figure 4b shows an example of the over-segmentation results.

Extracting a Coarse Building Region

A merging algorithm to merge separate segments of a single building is required to extract building regions based on the over-segmented results of the aerial images. Our merging method is based on the preliminary building regions and their heights obtained from the lidar data in previous stages. We define the *support ratio* as the ratio between the area of a particular segment that is supported by the lidar data to its total area as follows:

$$ratio = \frac{L_a \cap S_a}{S_a}, \quad (5)$$

where L_a is the area of the building region extracted from lidar data, and S_a is the area of a segment in aerial image. The *support ratio* naturally incorporates information extracted from the lidar data to represent the probability that a particular segment is a building region. If the ratio of a segment is more than a predefined threshold, then the segment becomes a building region. Finally, segments that are classified as building regions by common building regions in lidar data are merged into a group. In addition to the split and merge routine, color data from aerial images are used to remove tree segments directly connected to building regions, since they distract accurate building boundary extraction. More specifically, relatively small segments (in the image) in a building region (which is determined in the lidar data) that have high green value are removed. As shown in Figure 5, the partitioned building regions in the middle of Figure 5a are connected in Figure 5b by this split and merge method. Moreover, the majority of the tortuous building boundaries in Figure 5a are trimmed in Figure 5b, and tree segments directed connected to building regions are removed.

Extracting Precise Building Boundaries

Although it is possible to correct building regions from lidar data by the coarse building boundary process, the precise building boundaries cannot be obtained, since just trimming the tortuous coarse building boundaries may not guarantee the accuracy of extracting building boundaries process. The precise building boundaries are obtained in a closed-loop form by extracting edges from an aerial image, selecting the edge corresponding to the coarse building boundary, and conducting postprocessing by perceptual grouping such as edge linking and closed loop constructing.

Building Boundary Trimming

When applying edge information from the aerial image, it is difficult to use the coarse building boundary as the initial input. As shown in Figure 5b, extracted buildings have still irregular and tortuous boundaries. To extract precise building boundaries, a boundary trimming process that straightens the tortuous boundaries must be performed first. A

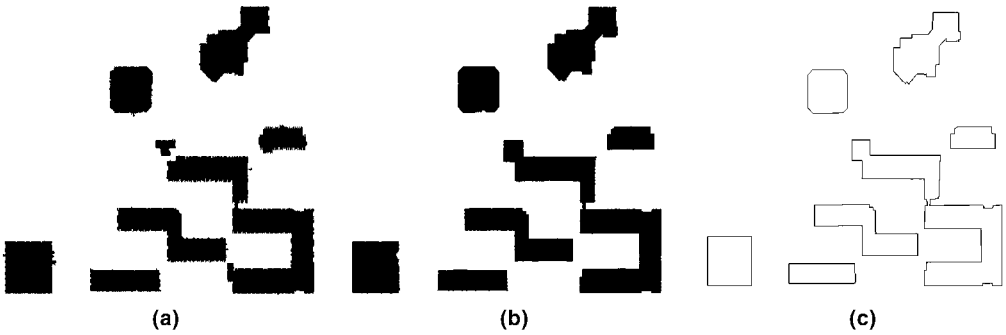


Figure 5. Extracted building region: (a) lidar result, (b) coarse building boundaries by split and merge, and (c) linearized building boundaries.

straight line boundary can be matched easily with an edge in the aerial image and consequently exert the characteristics of a man-made object composed of line features. In this paper, to find the corner points of a building boundary, the Douglas-Peuker method is applied. This method is a recursive split method to find the trimmed coarse building boundaries. Figure 5c shows the linearized building boundary of Figure 5b.

Extracting Building Boundary Candidates

As a feature in aerial image, the edge that is the best descriptor of the building is used. To extract the edge, the Nevatia-Babu edge detector is used. From detected edges, not all the edges are used. Only edges near a building boundary are necessary for the precise building boundary. Therefore, we can reduce the number of edges (building boundary candidates) by using information from the coarse building boundary and heights. Restricting the number of building boundary candidates reduces the computing cost and allows us to extract reliable building boundary candidates. Moreover, in this process, similar extracted edges can be grouped and modified according to the characteristics of man-made objects.

Searching for Edges Around a Building

An easy and simple method of restricting the number of boundary candidates is to find the edges near the building boundary. We accept edges within 20 pixels of a boundary as boundary candidates, considering registration error.

Height Compatibility

Height data from lidar measurements are used to eliminate some of the non-boundary edges around a building. If the registration of lidar data and photogrammetric imagery is correct, as shown in Figure 6, an edge in photogrammetric imagery is verified as a boundary candidate when its height is compatible with the average heights of the sampling window placed at either side of that edge. Note that the building inside region of a flat or sloping roof is higher than the building outside region of the roof boundary. Therefore, if an edge is part of a building boundary, then there is greater height difference between average heights of the two side sampling windows than the difference from bare-earth edge. In this paper, after extracting building candidate edges, binary values 0 and 255 have been assigned to building regions and bare-earth regions, respectively. If an edge is building boundary, then the difference of average value between left and right sampling window should be about 255. In other words, as shown in Figure 6, if an edge is a building boundary edge and if the average value of the left sampling window is 255 (bare earth region), then the average value of right sampling window should be 0 (building region), and vice versa. Using this attribute, we can select building boundary candidates from edges around a building.

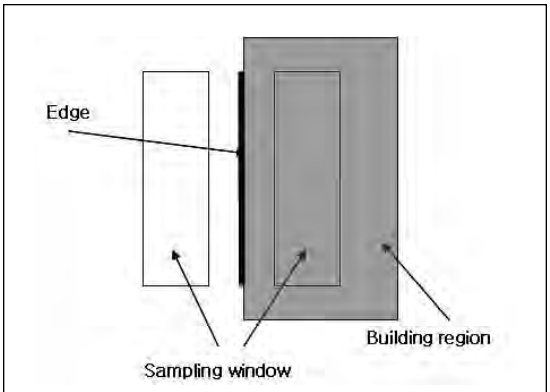


Figure 6. Building boundary candidate verification with height information.

Edge Linking

As a factor in man-made object features, a building boundary has generally straight line form. Although intuitively an edge is one long line segment, it can be shown that this edge is divided into short edges in the result of detected edges. Therefore, it is necessary to link short collinear edges that are within an allowed distance and have similar orientations. More specifically, two edges are linked if the closest distance between two edges is smaller than predefined threshold, and the angle between two edges is smaller than predefined threshold. Moreover, intuitively, the sum of two edge lengths should be smaller than the length of linked edge.

Constructing Precise Building Boundaries

Line Segments Matching

Using the extracted building boundary candidates described above, a boundary in the coarse building boundary result can be paired to several edges in the building boundary candidates. Several attributes exist for resolving matching ambiguities. The idea is to restrict the candidate matches to match one-to-one by some constraints, such as length ratio, angle, or

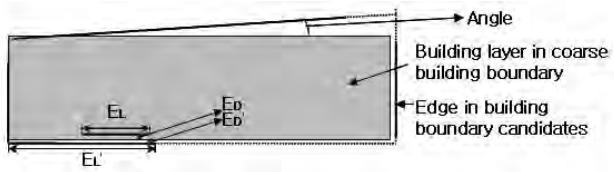


Figure 7. Selection of reliable corresponding edge.

distance between coarse building boundary and building boundary candidates. Because if we allow a 1:*n* relationship, shadow or street edges can disturb the performance of matching. Figure 7 shows an example of a building layer in coarse building boundaries and edges in building boundary candidates. Let E and E' be two edges in the building boundary candidates. Then, E_D and E_D' are the distances between a coarse building boundary and the building boundary candidates, and E_L and E_L' are the lengths of the building boundary candidates. In this specific figure, the angle θ of E and a matched coarse building boundary is 0 degrees.

For a given pair of line segments to be considered as a candidate match, constraints on the three attributes are applied to select the most consistent matches. First, the length ratio of a match can eliminate very short line segments that are false matches. Second, the distance measure can eliminate far-away edges. In this paper, we define the distance of two edges as the average distance between a point on one edge and a point on the other edge line. As shown in Figure 8, the distance between l_1 and l_2 can have two cases. Considering this, the distance is defined as follows:

$$E_d = ((\sum_i^n d_{li})/n + (\sum_j^m d_{2j})/m)/2, \tag{6}$$

where d_{1i} shown in right top case in Figure 8 is the distance between l_1 edge and a point on l_2 edge, while d_{2i} shown in the right bottom case is the distance between a point in l_2 edge and l_1. Due to distance error in the registration process mainly caused by point matching error, we allowed a maximum of 20 pixels distance between the coarse building boundary edge and building boundary candidate edges. Third, the similarity angle of two line segments, which is the most important attribute for defining similarity in relative position, is measured and pairs with angles over a threshold are discarded. We allowed a maximum of 15 degrees between two line segments. This process is applied to each building group. The combination of these constraints can be represented as follows:

$$Score = \begin{cases} w_1 L_R + w_2 \dfrac{15 - \theta}{15} + w_3 \dfrac{20 - E_D}{20}, & if\ \theta < 15\ and\ E_D < 20 \\ 0 & otherwise \end{cases}, \tag{7}$$

where θ denotes the angle between the two line segments, E_D denotes the distance between the two line segments, and w_1, w_2, w_3 denote weights of each attribute.

L_R is the ratio of the two lengths given by:

$$L_R = \begin{cases} \dfrac{C_L}{E_L} & if\ E_L > C_L \\ \dfrac{E_L}{C_L} & if\ C_L > E_L \end{cases}, \tag{8}$$

where C_L is the length of the coarse building boundary, and E_L is the length of the building boundary candidate. In our implementation, the weights $w_1 = 1$, $w_2 = 2$, $w_3 = 3$, and a threshold score of 3.0 is used.

It should be noted that no corresponding match exists when the score value is lower than the threshold. There are two strategies for this case. The first is to substitute an unmatched building boundary candidate with the corresponding coarse building boundary. This has the disadvantage of using inaccurate line segments but guarantees minimum accuracy. The second is to skip the unmatched building boundary. In other words, in this strategy, only matched edges are used to construct building boundaries. The advantage of this is that an edge in the building boundary candidates can compensate for a collapsed or false initial building boundary, but the accuracy of this strategy can be poor if all building boundary edges are not extracted from the aerial image. If the above constraint is fulfilled, we say that the pair of line segments considered is self-consistent and can form a precise building boundary. For each line segment in the coarse building boundary, we thus have a line segment match from building boundary candidates. Figure 9a shows the results of line segments matching.

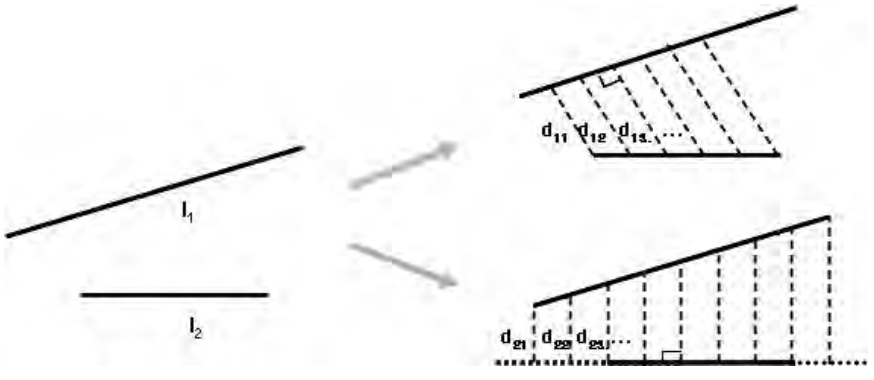


Figure 8. Distance between two edges.

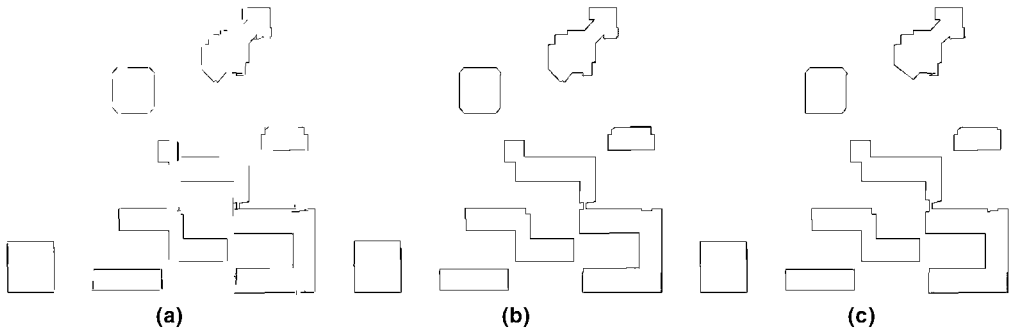


Figure 9. The result of precise building boundaries: (a) extracting corresponding edges, (b) substituting coarse building boundaries, and (c) using only edges in the aerial image.

Constructing a Closed Loop

The result shown in Figure 9a is not the complete building boundary. Since a building is a polyhedral structure of closed-loop form, the building boundaries in this figure should be closed polygons. To satisfy this condition, we must consider the missing features in the building sides and corners. Grouping only neighboring line segments is sufficient to compensate for the missing features. The reason is that the coarse building boundaries have the closed-loop form, and each matched line segment in the building boundary candidates has its one-to-one corresponding coarse building boundary.

Synthetically, we suggest two simple forms of closures. One is applied when the junction angle is more than 45 degrees, and the other is applied when the angle is less than 45 degrees. When the angle is more than 45 degrees, we can consider the intersection point of the extended lines of two edges as the junction point and the lines are extended to that point, respectively. However, when the angle is less than 45 degrees, the intersection point can be placed in wrong locations. Therefore, in this case, we just link the two edges. Moreover, we defined that the range of the junction angle is 0 to 90 degrees. Therefore, if the angle is 150 degrees, then it is considered as 30(180 to 150) degrees. Figure 10 shows examples of the two cases.

When the matching score of a coarse building boundary and a building boundary candidate is less than the threshold, there is no matching edge corresponding to the particular coarse building boundary. To handle this case, Figure 9b shows the result of substituting the coarse building boundary when there is no line segment matching pair, Figure 9c shows the result of consisting only matched edge. Generally, the result of 9b is better described than 9c, because all of the building boundary edges are not extracted due to similar textures at the building boundary. As shown in Figure 9b, three building boundaries that are at the top of the figure are described more specifically than Figure 9c. The matched edge and substituted coarse building boundary is linked as shown in Figure 10.

Experimental Results

Experiments designed to demonstrate the performance of our algorithm were carried out on two data sets. We show results for several examples in this section.

The Format of the Input Data

The lidar data and aerial images used in this paper, obtained airborne, represent a large 2 km $\times$ 2 km area of Daejeon in the Republic of Korea. The laser point density is about 3.2 pt/m^2. The resolution of the aerial images is about 5.2 pixels/m. Here, in the registration process, we scaled the lidar data result up to fit the high-resolution aerial images which utilizes all available image data resources.

Extracting Buildings from Lidar

Figure 11a shows building extraction results obtained from the proposed building candidate detection as previously described. This result consists of noise, small objects such as roadside trees, vegetation, and building regions. Note that we are able to

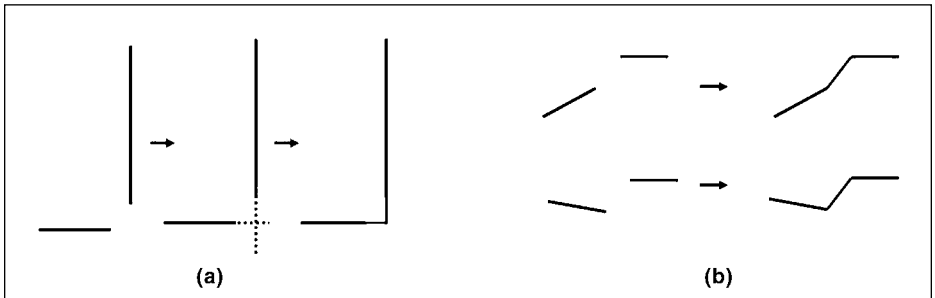


Figure 10. Closed loop forms: (a) when the angle of branches is more than 45-degrees, and (b) when the angle of branches is less than 45-degrees.

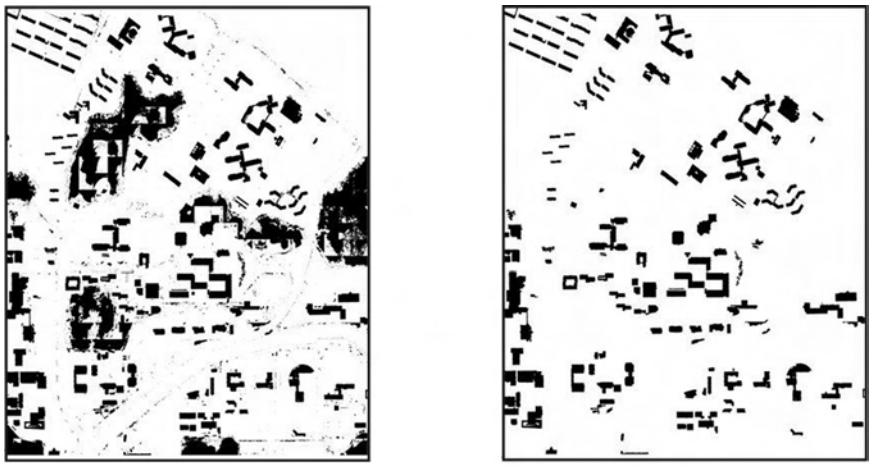


Figure 11. Extracted building region from lidar: (a) building candidates, and (b) after vegetation deletion.

intuitively identify each region by inspecting their boundary. Mainly quadrilateral-shaped regions can be classified as buildings, a few points or small regions can be classified as noise or small object regions, and irregularly shaped and relatively large regions are considered to be vegetation. As shown in Figure 11b and Figure 12, about 400 building candidates excluding noise or small regions are detected by the proposed method. The computed entropy value is the criterion to classify buildings and vegetation, where each region is classified as vegetation if the entropy is larger than threshold 2.5. As a result, 17 large groups were correctly classified as vegetation region from 400 building candidates. Figure 11b shows the result of deleting vegetation regions using both the proposed entropy based method and conventional area based method. In this figure, we can see that 17 of the large vegetation regions that have irregular boundaries and small regions including noise are removed correctly.

To quantitatively evaluate the classification results by the proposed method, we examine the ratio of the classified building, bare-earth, and vegetation regions that overlap with digital map data which is viewed as the ground truth (digital map). As shown in Table 1, the accuracy of the proposed method is 89.9 percent for building regions and 96.4 percent for bare-earth and vegetation regions. In this table, the number of each cell indicates the number of pixels in classified regions. The overall classification accuracy over the two classes, building and bare-earth including vegetation is 96.0 percent. In this analysis, the completeness for building is about 68 .1 percent for buildings. However, unfortunately, the lidar data was obtained about two years later than the digital map. During this period, new buildings have been constructed, and thus more buildings are appeared in lidar data than in the aerial image. As a result, the completeness is meaningless in our data.

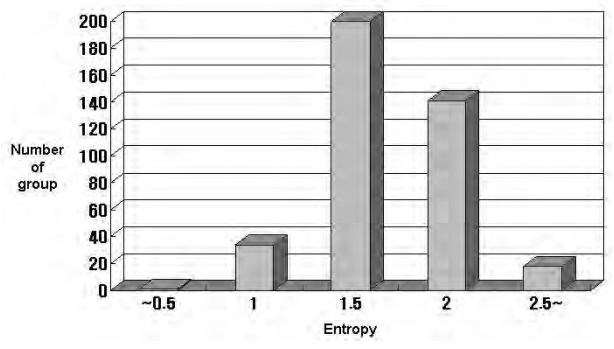


Figure 12. The number of group given entropy.

As another performance evaluation method, the number of detected buildings is compared to that of the digital map. More specifically, the ratio between the number of buildings, shown in both the digital map and the results of the proposed method, and the number of buildings in the digital map is computed. A total of 400 buildings are detected in the lidar data for the whole area, but only 54 buildings can be verified by the digital map which covers a smaller region, where the digital map shows 55 buildings. As a result, 98.2 percent of the buildings are detected correctly as described in Equation 9.

$$B_{DA} = \frac{NB_{Lidar \cap digitalmap}}{NB_{digitalmap}} = \frac{54}{55} = 98.2\%, \quad (9)$$

where B_{DA} denotes the building detection accuracy, $NB_{\text{Lidar}\cap\text{digitalmap}}$ denotes the number of buildings in both the lidar result and the digital map, and $NB_{\text{digitalmap}}$ denotes the number of buildings in the digital map.

Extraction of Building Boundary

Table 2 shows a list of threshold values used in our experiments. All the threshold values were determined heuristically. In Figure 11b, the region size to remove small areas

Table 1. Evaluation by Overlapped Pixel Number

Ground truth (digital map)	Classified as	
	Building	Bare-earth, vegetation
Building	57136(89.9%)	6431(10.1%)
Bare-earth, vegetation	26763(3.6%)	721305(96.4%)

TABLE 2. LIST OF THRESHOLD VALUES

The entropy classifying vegetation region	2.5
The region size to remove small areas in Figure 11b	25 pixels
The ratio to extract a coarse building boundary in the Extracting a Coarse Building Region section.	80%
The distance to search edge around building in the Searching for Edges Around a Building section.	20 pixels
The average distance in height compatibility in the Height Compatibility section	100
The angle for edge linking in the Edge Linking section	5°
The distance for edge linking in the Edge Linking section	50 pixels
The score of line segment matching in the Line Segments Matching section	3.0

was given as 25 pixels considering the size of a tree. In the Extraction of a Coarse Building Region section, the threshold 80 percent of the ratio is the value that is conventionally used. In the Extracting Building Boundary Candidates section, to search for candidate edges of building regions, we accepted edges within a 20 pixel distance from coarse building boundaries corresponding to the registration error, which is 20 pixels. In the *Height Compatibility* subsection, the average distance between left and right windows in height compatibility part should be 255 (white-black pixel value), but considering registration error, the distance threshold was given as 100. In the *Edge Linking* subsection, the angle and distance for edge linking were defined as 5° and 50 pixels, heuristically. In the *Line Segments Matching* subsection, the threshold score was given as 3.0 similar to the height compatibility case.

Analysis of Extracted Building Boundary

Unfortunately, it is difficult to acquire large data sets for a valid statistical evaluation. In addition, most of the building detection and description systems have different representational powers, and statistical evaluation on a small number of examples is less meaningful when the results depend on the choice of test data set. In this paper, we use two data sets to analyze the extracted building boundary. Set 1 shows a wide and complex building region, and Set 2 shows a region where buildings are described in detail and the building wall edges may interfere the line segment matching. Figure 13 shows the results of each process for the building boundaries in Set 2. Moreover, we compare the accuracy of building boundary extraction results obtained from only lidar data without any refinement, and lidar data with refinement, and from combining lidar data and photogrammetric imagery with intermediate coarse boundary extraction and precise refinement. The refinement process for the results extracted from the lidar data case is identical to the precise refinement of the combined case. To evaluate the results of our proposed method, ground truth of building boundaries are extracted manually from warped aerial images. For quantitative analysis of our result, we selected the data (aerial images) with higher accuracy to use as the ground truth. It is then possible to compare building extraction results with the ground truth. For performance evaluation, two kinds of methods are used. First, a buffer window is established around building boundaries corresponding to possible error. Figure 14 shows the buffer of Set 1, where the width of the buffer is 20 pixels. For evaluation, we check correspondence on each building boundary pixel, and define a ratio A_{bb} between the total

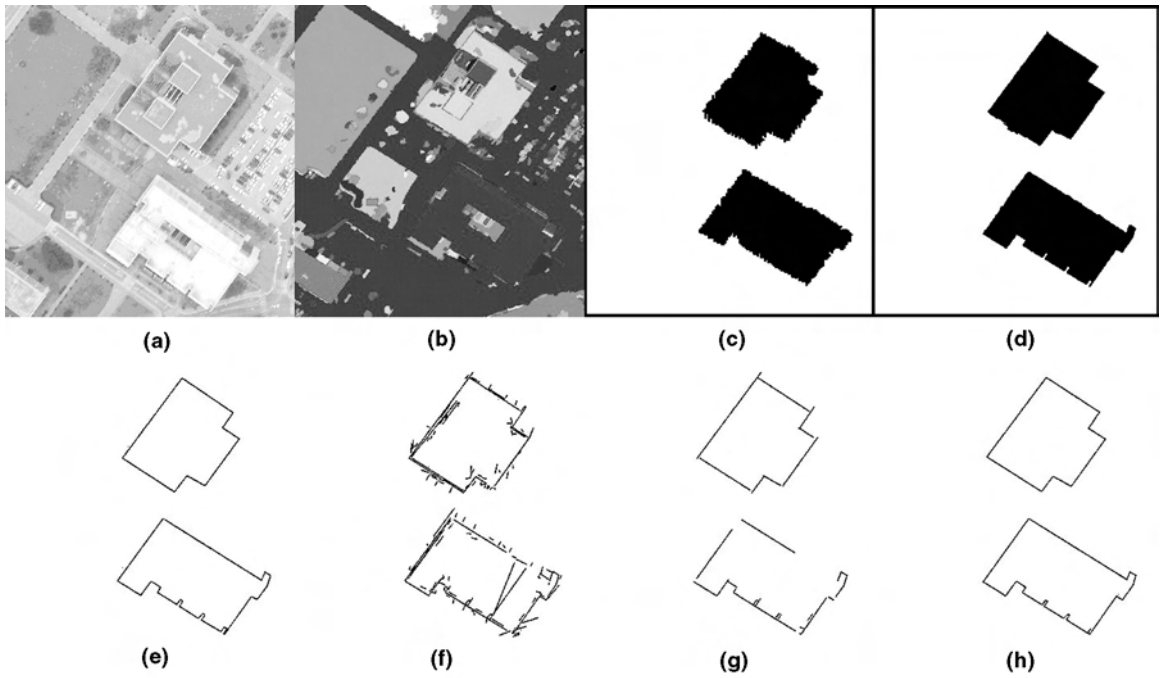


Figure 13. The results of extracting building boundaries in Set 2: (a) aerial image, (b) segmentation result of (a), (c) lidar result, (d) extracted coarse building regions by (b) and (c), (e) trimmed coarse building boundaries by Douglas-Peuker method, (f) building boundary candidates by constructing precise building boundaries, (g) the result of line segments matching, and (h) the result of constructing a closed loop. A color version of this figure is available at the ASPRS website: **www.asprs.org**.

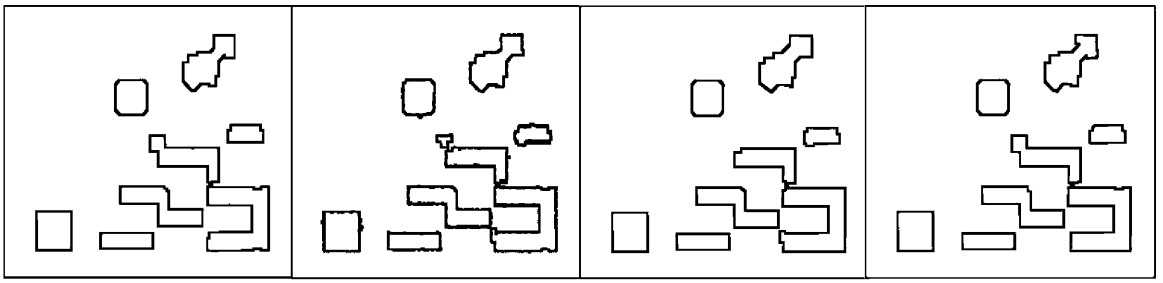


Figure 14. Buffers around the building boundries. (a) Ground truth of set 1, (b) Lidar result of set 1, (c) Building boundries by lidar result and precise method of set 1, (d) Building boundries by course and precise method of set 1.

number of buffer pixels N_{est} and the number of building boundary pixels within the buffer region N_{gt}:

$$A_{bb} = \frac{N_{gt \cap est}}{N_{est}}. \quad (10)$$

Second, Chamfer distance is applied to evaluate each method. To estimate Chamfer distance, the extracted building boundaries are converted from simple edges to a distance image. The result of each method is superimposed on the distance image of the ground truth, and the average of the distance values where the edges of the results intersect with the distance image represents the Chamfer distance. If the edges of a building boundary extraction results fits the distance image of the ground truth perfectly, the Chamfer distance is 0. In this paper, root mean square Chamfer distance is chosen:

$$Ch_{dts} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} v_i^2}, \quad (11)$$

where Ch_{dis} is Chamfer distance, v_i is distance value, and n is the number of points.

Table 3 and Table 4 show the accuracy of the three proposed methods. In Table 3, a performance improvement of about 2 to 11 percent was obtained in each step. In Table 4, because we scaled the lidar data result up to fit the high-resolution aerial images, a 5.2 pixel distance denotes 1 m. Here, a performance improvement of about 1 to 4 pixel distance was obtained in each step. However, the result of the coarse and precise building boundary in Set 2 did not show any improvement in the performance due to error in the segmentation result which is shown in the bottom building of Figure 13b. This false segmentation can occur where the texture of the building roof and walls are similar, and this has an effect on the Chamfer distance considerably. This is a limitation of applying the coarse building boundary extraction method based on color segmentation techniques exclusively, but as shown in Figure 13d, a more detailed description of building boundaries comparing to lidar results in Figure 13c is obtained by coarse boundary processes.

Table 3. Evaluation by Buffers Around Building Boundaries

Method	Data Set	
	Set 1	Set 2
Only lidar (%)	82.92	77.52
Lidar + refinement (%)	88.10	88.87
Lidar + image + refinement (%)	95.05	90.74

Table 4. Evaluation by Chamfer Distance

Method	Data Set	
	Set 1	Set 2
Only lidar (pixel distance (meters))	5.56(1.07m)	7.46(1.44m)
Lidar + refinement (pixel distance (meters))	4.55(0.88m)	3.24(0.62m)
Lidar + image + refinement (pixel distance (meters))	2.27(0.44m)	3.54(0.68m)

Conclusions

A new approach to extract the boundaries of complex buildings from lidar and photogrammetric imagery has been developed. The method is based on the application of information fusion to compensate for each sensor's shortcomings. To do this, we describe several techniques to group low-level features, such as height distributions, segments, and edges, into higher level features by using directional histograms, entropy, region segmentation and merging, line segments matching, and perceptual grouping. We used two kinds of measures to evaluate the proposed methods. From the evaluation, we have shown that the proposed multi-sensor fusion-based building detection method has improved the performance substantially in the accuracy of building boundaries compared to those of methods using lidar data only. The building boundary accuracy has improved more than 50 percent, and we could achieve satisfactory 3D reconstruction result based on extracted building boundary.

In our proposed method, many parts of proposed processes are performed separately, since each process in large area data requires high cost. However, in the very near future, the improvement of computing system will enable fast and incorporated processes within reasonable bounds. Therefore, we expect that the use of such methods will be automatically helpful in many other building detection and reconstruction problem domains as well. The problems of modeling complex buildings and automatic multisensor registration show many complexities requiring substantial future research, but we believe that this work indicates a promising approach.

Acknowledgments

This research was supported by the Defense Acquisition Program Administration and Agency for Defense Development, Korea, through the Image Information Research Center under the Contract UD070007AD.

References

Briese, C., N. Pfeifer, and P. Dorninger, 2002. Application of the robust interpolation for DTM determination, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXIV, pp. 55–61.

Brunn, A., and U. Weidner, 1997. Extracting buildings from digital surface models, *International Archives of Photogrammetry and Remote Sensing*, Vol. XXXII, pp. 27–34.

Comaniciu, D., and P. Meer, 2002. Mean shift: A robust approach toward feature space analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):1–18.

Habib, A., M. Ghanma, M. Morgan, and R. Al-Ruzouq, 2005. Photogrammetric and lidar data registration using linear features, *Photogrammetric Engineering & Remote Sensing*, 71(6):699–707.

Hartley, R., and A. Zisserman, 2003. *Multiple View Geometry in Computer Vision*, Cambridge University Press, Second edition, 89 p.

Kim, Z., and R. Nevatia, 2004. Automatic description of complex building from multiple images, *Computer Vision and Image Understanding*, Vol. 96, pp. 60–95.

Rottensteiner, F., and C. Briese, 2002. A new method for building extraction in urban areas from high-resolution lidar data, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXIV, pp. 295–301.

Rottensteiner, F., J. Trinder, S. Clode, and K. Kubik, 2004. Fusing airborne laser scanner data and aerial imagery for the automatic extraction of buildings in densely built-up areas, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Istanbul, Turkey, Vol. XXXV, Part B3, pp. 512–517.

Schenk, T., and B. Csatho, 2002. Fusion of lidar data and aerial imagery for a more complete surface description, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences Science*, Vol. XXXIV, pp. 301–317.

Sohn, G., and I. Dowman, 2003. Building extraction using LiDAR DEMs and IKONOS images, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Dresden, Germany, Vol. XXXIV, Part 3/W13.

Sun, J., Y. Lin, S.B. Kang, and H.Y. Shum, 2005. Symmetric stereo matching for occlusion handling, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 399–406.