Automatic Detection of Residential Buildings Using LIDAR Data and Multispectral Imagery

Mohammad Awrangjeb*, Mehdi Ravanbakhsh and Clive S. Fraser

Cooperative Research Centre for Spatial Information, The University of Melbourne
723 Swanston St, Carlton Vic 3053, Australia
Phone: +61 3 8344 9182, Fax: +61 3 9349 5185
Email: {mawr, m.ravanbakhsh, c.fraser}@unimelb.edu.au

Abstract

This paper presents an automatic building detection technique using LIDAR data and multispectral imagery. Two masks are obtained from the LIDAR data: a 'primary building mask' and a 'secondary building mask'. The primary building mask indicates the *void areas* where the laser does not reach below a certain height threshold. The secondary building mask indicates the *filled areas*, from where the laser reflects, above the same threshold. Line segments are extracted from around the void areas in the primary building mask. Line segments around trees are removed using the normalized difference vegetation index derived from the orthorectified multispectral images. The initial building positions are obtained based on the remaining line segments. The complete buildings are detected from their initial positions using the two masks and multispectral images in the YIQ colour system. It is experimentally shown that the proposed technique can successfully detect urban residential buildings, when assessed in terms of 15 indices including

*Corresponding author

completeness, *correctness* and *quality*.

Key words: Building detection, LIDAR, point cloud, multispectral, photogrammetry imagery, orthoimage, fusion, edge feature

---

1. Introduction

Building detection from remotely sensed data is important to the real estate industry, city planning, homeland security, disaster (flood or bush fire) management and many other applications. The automated extraction of building boundaries is also a crucial step towards generating city models (Cheng et al., 2008). Consequently, a large number of building detection techniques have been reported over the last few decades.

However, 100% successful automatic building detection is still an unrealized goal. There are several reasons to explain this situation (Sohn and Dowman, 2007). These include:

- *Scene complexity*: most of the scenes usually contain very rich information which provides a large number of cues with geometric or chromatic co-similarity to buildings, but belong to non-building objects.

- *Incomplete cue extraction*: there is always a significant loss of relevant building cues due to occlusions, poor contrast, shadows and disadvantageous image perspective.

- *Sensor dependency*: the primary data to support the building detection is available from a variety of sources with different resolution, each source having its own advantages and disadvantages for building detection.

Lee et al. (2008) have categorized building detection techniques into three groups. Firstly, there are many algorithms which use 2D or 3D information from photogrammetric imagery (Mayer, 1999). The complexity of separating buildings from other objects increases with the increase of image resolution as high-resolution images contain more detailed information (Cheng et al., 2008), along with occlusions and shadows (Yong and Huayi, 2008). The derivation of 3D information, for example, the depth information from stereo by multiple images (Sun et al., 2005), is even more complicated (Vu et al., 2009). In addition, nearby trees of similar height also make the use of such derived range data difficult (Lee et al., 2008).

Secondly, there have been several attempts to detect building regions from LIDAR (LIght Detection And Ranging) data. This task has been largely solved by classifying the LIDAR points according to whether they belong to bare-earth, buildings, or other object classes (Lee et al., 2008). In fact, the introduction of LIDAR has offered a favourable option for improving the level of automation in the building detection process when compared to image-based detection (Vu et al., 2009). Oude Elberink (2008) has discussed a number of problems with building detection using LIDAR data and it has been shown that the use of raw or interpolated data can influence the detection performance (Demir et al., 2009). Moreover, there may be poor horizontal accuracy for building edges (Yong and Huayi, 2008) and it is hard to obtain a detailed and geometrically precise boundary using only LIDAR point clouds (Cheng et al., 2008). The quality of regularized building boundaries also depends on LIDAR resolution (Sampath and Shan, 2007).

LIDAR and photogrammetric imagery each have particular advantages

and disadvantages in horizontal and vertical positioning accuracy. Compared with photogrammetric imagery, LIDAR generally provides more accurate height information but less accurate boundary lines. Unfortunately, some regions in LIDAR data have null values due to self-occlusion of a building or if they contain water. Photogrammetric imagery can provide extensive 2D information such as high-resolution texture and colour information as well as 3D information from stereo images. As a result, several authors have promoted an integration of LIDAR data and imagery as a means of advancing building detection (Rottensteiner et al., 2005; Yong and Huayi, 2008; Cheng et al., 2008; Demir et al., 2009).

The third category of methods does use both LIDAR data and photogrammetric imagery. More specifically, intensity and height information in LIDAR data can be used with texture and region boundary information in aerial imagery to improve accuracy (Lee et al., 2008).

However, the question of how to integrate the two data sources for building boundary extraction still arises; few approaches with technical details have thus far been published (Rottensteiner et al., 2005). The question of how to combine the two different data sources in an optimal way so that their weaknesses can be compensated effectively is an active area of current research (Yong and Huayi, 2008).

Regarding performance evaluation, there is a current lack of uniform and rigorous evaluation systems, and an absence of standards (Rutzinger et al., 2009). Indeed, evaluation results are often missing from published accounts of building detection (Yong and Huayi, 2008); the use of 1 to 2 evaluation indices only has characterized many studies (Demir et al., 2009; Vu et al.,

2009).

This paper aims at following two goals: a successful integration of the LIDAR data and photogrammetric imagery for building detection so that the increased detection performance is obtained and development of an automatic performance evaluation system using 15 evaluation indices.

The proposed automatic building detection technique uses raw LIDAR data and orthoimagery. Two masks are obtained from the LIDAR data: a 'primary building mask' and an 'secondary building mask'. Line segments around the black shapes (absence of height data) in the primary building mask constitute the initial building positions. The final buildings are then detected extending their initial positions using the multispectral images, transformed into the YIQ (intensity, hue and saturation) colour system. The two masks ensure accurate delineation of the buildings. In particular, the primary building mask helps separate detected buildings when they are very close to each other and the secondary building mask helps to avoid extensions to initial positions outside a building when the roof and ground have similar colour information. It is experimentally shown that the proposed technique can detect rectilinear buildings with a favourable success rate, especially within the Australian urban environment for which it was primarily developed.

The proposed detection technique has similarities to that reported by Sohn and Dowman (2007) and Cheng et al. (2008) in the sense that it uses line segments and a regularization step (adjustment) employing dominant line angles.

The proposed automatic evaluation system uses both object- and pixel-

based indices. In addition, though the pixel-based evaluation indirectly reflects the horizontal accuracy, the geometric evaluation is introduced as a means of direct estimation of the horizontal, geometric or positional accuracy.

The performance of the proposed building detection approach has been evaluated through using 15 indices in three categories, these being object-based, pixel-based and geometric. Most of the indices have been adopted from the literature and the remainder are proposed here for a more complete evaluation.

The rest of the paper is organized as follows: Section 2 presents a review of both existing integration techniques for photogrammetric imagery and LIDAR data for both building detection and performance evaluation systems. Section 3 details the proposed building detection technique. The proposed evaluation system with experimental test results is discussed in Section 4. Finally, concluding remarks are offered in Section 5.

2. Related Work

2.1. Integration of LIDAR and Imagery

Building detection techniques integrating LIDAR data and imagery can be divided into two groups. Firstly, there are techniques which use the LIDAR data as the primary cue for building detection and employ the imagery only to remove vegetation (Rottensteiner et al., 2005; Vu et al., 2009). As a result, they suffer from poor horizontal accuracy for the detected buildings. Rottensteiner et al. (2005) employed the Dempster-Shafer theory as a data fusion framework to classify points as buildings, trees, grassland or bare soil.

However, the detection performance was adversely affected for small buildings (Rottensteiner et al., 2007). The reason is that if the Dempster-Shafer model is not properly trained, then the miss-classification rate increases considerably (Khoshelham et al., 2008). Vu et al. (2009) used a morphological scale-space for extracting building footprints from the elevation data and then removed vegetation areas using the spectral data. The detection performance was low and high computational complexity was reported because of using the scale-space.

Secondly, there are integration techniques (Haala and Brenner, 1999; Chen et al., 2004; Sohn and Dowman, 2007; Lee et al., 2008; Demir et al., 2009) which use both the LIDAR data and the imagery as the primary cues to delineate building outlines. They also employ the imagery to remove vegetation. Consequently, they offer better horizontal accuracy for the detected buildings. The proposed building detection technique falls into this group. Haala and Brenner (1999) applied a pixel-based classification where the normalized DSM (nDSM) was used as an additional channel to the three spectral bands of the aerial imagery. Chen et al. (2004) followed a region-based segmentation of nDSM and orthoimages and then used a knowledge-based classification to detect building. However, this method did not show how to cope with erroneous lines (Sohn and Dowman, 2007) and could not detect small buildings.

Sohn and Dowman (2007) employed a data-driven approach on the optical imagery and a model-driven approach on the point cloud to extract rectilinear lines around buildings. Extracted lines were regularized by analyzing the dominant line angles. Cheng et al. (2008) proposed a similar technique with

precise geometric position. Lee et al. (2008) extracted the initial building boundaries from the LIDAR data and then enhanced the initial boundaries using colour information, after which edge matching and perceptual grouping techniques were applied to yield the final building boundaries. Demir et al. (2009) applied four different methods to achieve an improvement by combining the advantages and disadvantages of these approaches and used the edge information from images for quality improvement of the detected buildings.

2.2. Evaluation Systems

Performance evaluation systems reported in the literature can be divided into two groups: those using overlapping thresholds (Rottensteiner et al., 2005; Rutzinger et al., 2009; Lee et al., 2008) and those not using any thresholds (Shan and Lee, 2005; Shufelt, 1999). Threshold-based systems use one or more overlapping thresholds while making correspondences between detected and reference building sets. The problem with threshold-based systems is that they are subjective and likely to be controversial since there is no unique way to select the thresholds (Shufelt, 1999).

The evaluation systems can also be categorized into pixel-based systems (Rottensteiner et al., 2005; Rutzinger et al., 2009; Lee et al., 2008) and object-based systems (Rutzinger et al., 2009). While the latter counts the number of buildings and offers a quick assessment, the former is based on the number of pixels and provides more rigorous evaluation (Song and Haithcoat, 2005). The pixel-based evaluation indirectly corresponds to the horizontal accuracy of the detected building footprints.

In Rottensteiner et al. (2005) and Rutzinger et al. (2009), a correspon-

dence was established between a detected building and a reference building if they overlapped each other either strongly, more than 80% overlap, or partially, 50% to 80% overlap. Both of the above evaluation systems do not reflect the actual detection scenario. Firstly, the presence of *false positive* and *false negative* detections is not considered at all. Secondly, there may be many-to-many relationships between the detected and reference sets and such relationships are considered as error (Shan and Lee, 2005). Finally, merging and splitting of the detected buildings as in Rutzinger et al. (2009) does not necessarily correspond to the actual performance.

Without using a particular overlapping threshold, Shufelt (1999) showed the detection performance graphically as the overlapped area varied from 0-100%. Shan and Lee (2005) presented results by histograms showing the frequency of buildings as functions of *underlap*, *overlap*, *extralap*, *crosslap*, and *fitness*. The number of *false negative* buildings was indicated by the frequency at 100% *underlap* and the number of *false positive* buildings was indicated by the frequency both at *crosslap* 0 and 0% *fitness*.

3. Proposed Building Detection Technique

The proposed automatic building detection technique uses LIDAR data and colour orthoimagery. It has four major steps. Firstly, two masks, a 'primary building mask' and an 'secondary building mask', are generated from the LIDAR data. The primary building mask indicates the *void areas* where there are no laser returns below a certain height threshold. The secondary building mask indicates the *filled areas*, from where returns indicate an elevated object above the same height threshold. Secondly, line segments

from around the void areas in the primary building mask are extracted. Line segments around trees are removed using Normalized Difference Vegetation Index (NDVI) values derived from the multispectral images. Thirdly, initial building positions are recovered based on the remaining line segments. Finally, the complete building footprints are obtained from their initial positions using the two masks and the orthoimagery in the YIQ colour system.

3.1. Overview

Fig. 1 shows the flow diagram of the proposed building detection technique. The input information consists of a LIDAR point cloud, a DEM (digital elevation model) and multispectral orthoimagery. The point cloud and orthoimagery are registered to each other before being used as inputs. The primary and secondary building masks are first derived from the LIDAR data, along with NDVI values from the orthoimagery. The initial building positions are derived from the primary building mask. The colour information in the multispectral images is usually in the RGB system and therefore is converted into the YIQ system. The final buildings are obtained by extending their initial positions using the two masks and the YIQ colour information.

The following subsections detail the proposed detection technique. Section 4.4.1 presents the sensitivity analysis of important parameters used by the detection algorithm.

3.2. Height Threshold and Masks

While the primary building mask M_p is used for obtaining the initial building positions as rectangular areas, the secondary building mask M_s is

used as an indication of a maximum building size around an initial building position during the detection of a final building from its initial position.

All pixels in M_p are initially assigned 0 (false), but in M_s are assigned 1 (true). The two masks are derived simultaneously by first dividing M_p (and hence M_s) into tiles of size 450×450 image pixels, since there may be different representative ground heights H_g in a large area. H_g is calculated separately for each tile from the corresponding DEM data. The DEM and LIDAR data are also grouped following the tiles of M_p. For each tile, H_g is simply estimated as the average of the height data from the corresponding DEM. Fig. 2(a)-(b) shows the tiles of masks on an orthoimage and the groups of LIDAR data.

In order to obtain the masks for each tile, a threshold $T_h = H_g + 2.5m$ is applied. If the LIDAR height of a point (x, y) is less than T_h, the corresponding pixel in M_p is assigned 1. If the height is greater than T_h, the corresponding pixel in M_s is assigned 0. In addition, since the horizontal resolution of LIDAR data is generally lower than that of the orthoimage, all the pixels in a 5×5 neighbourhood of (x, y) are also assigned 1 for M_p or 0 for M_s. The size of the neighbourhood can be adjusted based on the relative resolutions of the LIDAR data and the orthoimage.

From Fig. 2(c) it can be seen that the majority of the buildings are distinguishable in the primary building mask. However, when buildings are very close to each other, many are not clearly distinguishable in the secondary building mask, as shown in Fig. 2(d). Colour information from the orthoimagery is therefore used for more accurate detection of the buildings (Section 3.4).

3.3. Initial Building Positions

Initial building positions $B_{ini} = \{b_{ini,i}\}$, $1 \leq i \leq m$, where m is the number of detected positions, are detected as rectangular patches from the primary building mask M_p, with the black areas in Fig. 2(c) being the initial building positions. This section describes how each of those areas is detected as a rectangle or as a combination of two or more rectangles.

Three steps are followed to obtain B_{ini} from M_p. Firstly, lines around the black shapes from M_p are formed. Secondly, the lines are adjusted and extended. Finally, rectangular shapes are obtained using these lines.

3.3.1. Line Detection

The Canny edge detector (Canny, 1986) is first used to find all the edges in M_p and then the short edges are discarded. Edges of less than $3m$ (20 pixels) in length are considered short, assuming that the minimum building length or width is $3m$.

Since there may be noise and local variation introduced by the neighbourhood filling technique during the mask generation phase, a Gaussian kernel with scale $\sigma = 3$ is utilized to smooth each edge.

Corners (absolute curvature maxima points) are then detected on each of the smoothed curves using a fast corner detector described in Awrangjeb et al. (2009). The smoothed curves are then decomposed into line segments. On each edge, all the pixels between two corners or a corner and an endpoint or two endpoints when enough corners are not available, are considered as separate line segments. Again, short line segments, whose lengths are less than $3m$, are discarded.

The detected corners and edge endpoints may not be well localized to the building corners. In order to align the detected line segments with the building edges, a least-squares straight-line fitting technique is applied. With each line segment a point P_{in} is recorded. This 'inside-point' indicates on which side of the line the building is recorded.

Obviously, some line segments around trees are obtained and in order to avoid further processing of these, a rectangle of $3m$ *width* on the building side is formed. The sigma of the NDVI value Υ inside the rectangle is then employed, such that for a line segment if the mean of Υ is above a threshold $T_{ndvi} = 48$, the line segment is classed as a tree-edge and removed. After the application of the NDVI threshold on the extracted lines, the removed lines are shown in Fig. 3(a) with circles at their centres.

It was found that the NDVI did not have a high discriminating power. Many of the tree-edges could not be removed using T_{ndvi}. If a low T_{ndvi} value was applied these tree-edges could be removed, but many important line segments which indicate the initial positions of buildings were also removed. A similar effect was reported by Rottensteiner et al. (2007) who applied a post-classification technique to improve the performance. Consequently, in this investigation a minimum building length threshold ($3m$) has been applied to the extracted line segments to remove small vegetation areas. In addition, assuming that the buildings and their sides are locally parallel or perpendicular to each other, the line segments are adjusted as discussed below. This adjustment procedure removes a tree-edge which is neither locally parallel nor perpendicular when compared to its neighbouring line segments.

3.3.2. Adjusting and Extending Lines

Under the assumption that the longer lines are more likely to be building edges, the extracted lines are sorted based on their lengths. Then, in an iterative procedure starting from the longest line l_i and taking it as a reference, the angle between the reference l_i and each line l_j in its neighbourhood is estimated. A circular neighbourhood around the centre of l_i is then considered. The radius of this neighbourhood is set as the maximum building length, $50m$ in this investigation. If l_i and l_j are either parallel or perpendicular to each other, to within a $\frac{\pi}{8}$ angular difference, the rotation angle θ_r for l_j is estimated.

There may be buildings of different orientations in an area. This means that one building or a group of buildings may have a different orientation when compared to others in the neighbourhood. In order to avoid wrong adjustments of the extracted lines, the lowest rotation angle θ_r is recorded for each l_j over all iterations.

The above iterative procedure may be optionally terminated after a significant number of iterations, say 50% of the number of total extracted lines. After the iterative procedure, each l_j and its P_{in} are rotated with respect to the line centre by its recorded angle θ_r. If a rotation angle is not recorded for l_j, then this l_j is removed as a tree-edge. After the above adjustment procedure, the removed lines are shown in Fig. 3(b) with circles at their centres.

Each of the adjusted line segments may not represent a complete side of a building. The line may be disrupted by trees, noise introduced in the edge detection process, and by the neighbourhood filling effect. Therefore,

both ends of each adjusted line C_1C_2 are extended by considering a rectangle of *length* $L_e = 3m$ and *width* $W_e = 3m$ on each side (see Fig. 4(a) for C_2). Inside the rectangle the percentage of black pixels in M_p, Ω, should be high and more than 70%, and the mean Υ should be low at less than 48. This extension process continues iteratively and if any of the conditions fail at any iteration, W_e is halved. The process stops when W_e is less than the successive LIDAR point distance (i.e., $\leq 0.4m$ in this case).

3.3.3. Initial Buildings

Since long line segments represent more accurate building edges than short ones, the extended line segments are sorted again in descending order based on their lengths. In an iterative procedure, an initial building position is detected using the first longest line segment, another using the second longest line segment and so on. The rectangular positions are recorded in a set $B_{ini} = b_{ini,i}$, where $0 \leq i \leq m$, of four-points, one for each corner of a rectangle. B_{ini} is initially empty. Before detecting a rectangle using a line segment C_1C_2 in each iteration, C_1C_2 is tested to ascertain whether it is already in a detected rectangle $b_{ini,i}$.

In order to detect a new rectangle using C_1C_2, an initial rectangle C_1C_2NM, with *length* $L_b = |C_1C_2|$ and *width* $W_b = 1.5m$, is formed on the building side. Then three sides MN, C_1M and C_2N of C_1C_2NM are extended outwards with respect to P_{in} (Fig. 4(b)) using the same technique as that applied to extend the extracted lines, as discussed above.

After extension of three sides, if any of the sides of C_1C_2NM is not at least $3m$, C_1C_2 is removed as a tree-edge. Fig. 3(c) shows the initial building positions.

3.4. Final Building Positions

The final building positions are obtained from their initial positions by extending each of the four sides. Image colour information and the two masks M_p and M_s are considered during the extension. The colour information is basically used to extend the initial positions; M_p is used to avoid unexpected extension of an initial position over more than one actual buildings, and M_s is used to avoid unexpected extension of an initial position beyond the actual building roof.

In practice, there are different shapes of rectilinear buildings. We have adopted a definition whereby a simple rectangular building (or building-part) is termed an 'I' shape. Two adjoining perpendicular 'I' shapes then form either an 'L' or 'T' shape building, whereas three connected rectangular building parts form a 'U' shape, and four connected rectangular parts around an open central area are termed a 'C' shape building.

If there are different rectangular initial positions for the same building, it could be for one of the following two reasons. Firstly, an 'I' shape building may be detected more than once. Secondly, the building is 'L', 'T', 'U' or 'C' shaped. In both of the above cases, the initial positions may overlap partly or fully before or after their extensions. While in the first case, an overlap is unexpected and has a negative impact in the detection performance, in the second case an overlap is considered as a *natural overlap* and is expected to join different detected parts of the same building, if necessary in any later applications.

However, it is hard to decide which overlap is unexpected and which is natural. If an initial building is completely within an already extended

building or building part, it is removed assuming that it is an unexpected overlap. Otherwise, it is extended assuming that it is a natural overlap. Before extending initial positions to obtain final positions, a preprocessing step is executed.

3.4.1. Preprocessing Initial Buildings

An initial building position may go outside the actual building roof due to a misregistration between the orthoimage and the LIDAR data. In order to avoid this, since the initial position will be extended outwards while obtaining the final position, its *length* and *width* are reduced by 15% before extension. For each reduced building position $ABCD$, the dominant colour threshold pairs $T_Y = [l_Y, h_Y]$, $T_I = [l_I, h_I]$ and $T_Q = [l_Q, h_Q]$ are estimated for intensity Y, hue I and saturation Q, respectively. Each dominant colour threshold pair indicates a range denoted by its *low* l and *high* h values.

In order to find threshold pairs for each band ($Y \in [0, 1]$, $I \in [-0.5957, 0.5957]$ and $Q \in [-0.5226, 0.5226]$), a histogram is generated for its values within $ABCD$, over 10 bins for Y or 20 bins for I and Q. Fig. 5 shows different types of histograms. The uphill and downhill histograms in Figs. 5(a)-(b) are two basic histograms and, practically, a histogram of Y, I or Q is a combination of these two. Fig. 5(c) shows one of the simple combinations that occurred most frequently in the experiments conducted for this investigation.

The histogram is divided into n parts, where $n \geq 1$. The value of n is usually 1, but may be greater than 1 if more colours appear on the building roof. Each part is from a minimum frequency bin to the next minimum frequency bin, or from a minimum frequency bin to the next maximum frequency bin, or from a maximum frequency bin to the next minimum frequency bin, if

enough neighbouring minimum frequency bins are not available. For example, the hue histogram in Fig. 5(c) is divided into two parts. Part 1 is between Min 1 to Min 2, but Part 2 is from Min 2 to Max 2 since there is no Min 3.

The histogram parts are sorted in descending order based on their total number of points, or total frequencies, t_i, where $1 \leq i \leq n$. Starting from the part that has the highest total points $max(t_i)$, points are accumulated for each part from its maximum bin towards its minimum bins, adding the next largest bin at a time. The accumulation stops for a part if the number of accumulated points is at least 97% of this part and *low* and *high* threshold values are recorded at stop positions. In this way threshold pairs are estimated for other parts iteratively and the iteration terminates if the total number of points of the already used parts is at least 90% of $ABCD$. This means parts having very low t_i values are not considered, which helps to avoid extension of an initial position towards a vegetation area whose small region is on the roof and within the initial position, but the major region is outside the building.

3.4.2. Extending Initial Positions

The initial building positions $B_{ini} = \{b_{ini,i}\}$ are sorted in descending order of their *length* or *area*, since both of these sorted lists were found to offer the same performance. Then in order to obtain final building positions $B_{fin} = \{b_{fin,i}\}$, all initial positions are extended one after another, starting from the one having the longest *length* or largest *area*.

To extend an initial position $b_{ini,i}$ denoted by a rectangle $ABCD$, its four sides are extended separately. To extend a side, say AB, a rectangle

$ABNM$, with *length* $L_f = |AB|$ and *width* $W_f = 0.35m$, opposite to P_{in}, is considered. For $ABNM$ the percentages of Y, I and Q within threshold pairs $T_Y = [l_Y, h_Y]$, $T_I = [l_I, h_I]$ and $T_Q = [l_Q, h_Q]$, respectively, are computed. Let these percentages be λ, χ and μ. The percentages of black pixels in the primary and secondary building masks for $ABNM$ are also computed. Let these percentages be ς and ν. If λ, χ and μ are above 40% and ν is above 90%, AB is extended by replacing M by A and N by B. This extension procedure of AB continues iteratively and in each iteration the value of ς is checked and it should either be the same as or less than in the previous iteration. If ς becomes below 10% there is a high probability that the extension procedure will soon end. However, if ς starts increasing thereafter it is the position where AB is being extended over a neighbouring object, either a building or a tree. If this is the case, the extension procedure for AB immediately terminates. Otherwise, if any other condition fails, for example, if any of λ, χ and μ is below 40% or ν is below 90%, W_f is divided by 2 and the extension of AB continues. The procedure finally terminates if W_f is less than the image ground resolution (i.e., $< 0.1m$ in this investigation). After extension of all four sides of $ABCD$, the extended rectangle is obtained. Fig. 6 shows the final detected buildings in four tested scenes.

4. Performance Evaluation

The proposed threshold-free evaluation system makes *one-to-one* correspondences using nearest centre distances between detected and reference buildings. The reference buildings are obtained using manual measurement from the orthoimagery (Section 4.1). Altogether 15 indices are used in three

categories (object-based, pixel-based and geometric) to evaluate the performance. Most of these have been adopted from the literature and the rest are proposed for a more complete evaluation (Section 4.2). Section 4.3 details the experimented data sets and Section 4.4 presents results and a discussion.

4.1. Evaluation System

For evaluation, two sets of data were used, in which each building is represented either as a rectangular entity, for 'I' shape building, or a set of rectangular entities, for 'L', 'U' and 'C' shapes. The first set $B_d = \{b_{d,i}\}$, where $0 \leq i \leq m$ and m is the number of detected rectangular entities, is known as the *detected set*. It is obtained from the proposed automatic building detection technique. Each entity $b_{d,i}$ is an array of four vertices and the centre (intersection of two diagonals) of a rectangular detected entity. The second set $B_r = \{b_{r,j}\}$, where $0 \leq j \leq n$ and n is the number of reference entities, is termed the *reference set*. It is obtained from manual building measurement within the orthoimagery. Each entity $b_{r,j}$ is an array of four vertices and the centre of the rectangular reference entity.

To find the reference set B_r, manual image measurement is used. Any building-like objects above the height threshold T_h (Section 3.2) are included in B_r. As a result some garages (car-ports) whose heights are above T_h are also included, but some building parts (verandas) whose heights are below T_h are excluded. Different building parts are referred to separate rectangular entities. Consequently, there is one entity for 'I' shape, two entities for 'L' shape, three entities for 'U' shape, four entities for 'C' shape and so on.

It is natural that different rectangular entities of the same building overlap each other. In B_r, two overlapping entities must always belong to the

same building and represent two connected building parts (Fig. 7(a)). Such an overlap is defined as a *natural overlap* and for identification purposes a *building identification number* b_{id} is assigned to each reference entity, this being stored in $b_{r,j}$, in addition to the four vertices. Entities of the same building are assigned the same b_{id}, but those of the different buildings are assigned different b_{id} values.

In B_d, the situation is different. Here two overlapping entities may belong to the same building and represent two connected building parts. In such a case, this overlap is a *natural overlap* (Fig. 7(a)) and it is not counted as an error in the proposed evaluation. In all other cases, the overlap is counted as an error in the evaluation system. For example, the overlapping entities may represent the same building (multiple detection, Fig. 7(b)) or constitute combinations of true and false detections (Figs. 7(c)-(e)).

In an approach similar to that of Song and Haithcoat (2005), a detected entity is counted as correct if any of its part overlaps a reference entity. However, unlike existing evaluation systems (Rottensteiner et al., 2005; Rutzinger et al., 2009), a *pseudo one-to-one* correspondence is established between the detected and reference sets without using any thresholds. *Pseudo one-to-one* correspondence means that each entity in one set has at most one correspondence in the other set. If a detected entity overlaps only one reference entity which is not overlapped by any other detected entity, then a true correspondence is established between them. If a detected entity overlaps more than one reference entity, then the nearest reference entity (based on the distance between centres) is considered as a true correspondence for the detected entity. The same rule is applied when a reference entity is overlapped by more

than one detected entity. As a consequence, there will be no correspondence for *false positive* and *false negative* entities.

4.2. Evaluation Indices

Altogether, 15 performance evaluation indices in three categories have been adopted: object-based evaluation, area- or pixel-based evaluation and geometric evaluation. For pixel-based evaluation, pixels in the orthoimage are used for all detected and reference entities. The geometric evaluation is separated from the other two as such a evaluation estimates the positional accuracy and counts neither the number of objects nor the number of pixels. In the following subsections, different indices in the three categories are discussed. Note that the definitions of *true positive* (TP), *true negative* (TN), *false positive* (FP) and *false negative* (FN) have been adopted from Lee et al. (2003). In addition, a new term *multiple detection* (MD), which indicates that for an entity presented in the reference set there are two or more entities in the detected set, has also been used.

4.2.1. Object-based Indices

The following seven indices are used for object-based evaluation to evaluate the number of buildings counted. Completeness C_m, also known as *detection rate* (Song and Haithcoat, 2005) or *producer's accuracy* (Foody, 2002), *correctness* C_r, also known as *user's accuracy* (Foody, 2002) and *quality* Q_l have been adopted from Rutzinger et al. (2009). The remaining four are defined as:

1. *Multiple detection rate* is the percentage of multiply and correctly detected entities in the detected set. As shown in Fig. 7(b), a building or

a building part can be detected more than once and all these detected entities correspond to a single entity in the reference set. The closest detected entity with respect to the reference entity is marked as a TP and all others as MDs. The *multiple detection rate* is defined as

$$M_d = \frac{|MD|}{|TP| + |FP| + |MD|}, \tag{1}$$

where $|.|$ denotes the set cardinality. Note that $|TP| + |FP| + |MD|$ denotes the total number of entities in the detected set.

2. *Detection overlap rate* is the percentage of overlap in the detected set. It is defined as

$$D_o = \frac{O_d}{|TP| + |FP| + |MD|}, \tag{2}$$

where O_d is the number of detected entities that overlap other detected entities. However, the natural overlaps are excluded.

3. *Detection cross-lap rate* is defined as the percentage of detected entities which overlap more than one reference entities and expressed as:

$$C_{rd} = \frac{C_{ld}}{|TP| + |FP| + |MD|}, \tag{3}$$

where C_{ld} is the number of detected entities which overlap more than one reference entity and the natural overlaps are again excluded.

4. *Reference cross-lap rate* is defined as the percentage of reference entities which are overlapped by more than one detected entity and this is expressed as

$$C_{rr} = \frac{C_{lr}}{|TP| + |FN|}, \tag{4}$$

where C_{lr} is the number of reference entities which are overlapped by multiple detected entities, with the natural overlaps being excluded.

A good building detection system should have high C_m and C_r values, but low M_d, D_o, C_{rd} and C_{rr} values, while $1 - C_r$ indicates the false alarm rate of the system. Q_l makes a compromise between C_m and C_r (Heipke et al., 1997).

4.2.2. Pixel-based Indices

For area- or pixel-based evaluation, pixels in the orthoimage are used for all detected and reference entities. For an FP detected entity, all the pixels within it are FP_p pixels (subscript p stands for pixels). For an FN reference entity, all the pixels within it are FN_p pixels. For a TP detected entity, there are two types of pixels: all the pixels within it that also appear in the corresponding reference TP entity are TP_p pixels and the rest are FP_p pixels. Similarly, for a TP reference entity, there are two types of pixels. All the pixels within it that also appear in the corresponding detected TP entity are TP_p pixels (counted only once) and the rest are FN_p pixels with all other pixels being TN_p pixels. Note that within the natural overlapping area the pixels are counted only once though they may be detected twice (as they are common to two detected entities on the same building). MDs are not considered in the pixel-based evaluation.

A total of 7 pixel-based evaluation indices are used, these being: *completeness* C_{mp}, also known as *matched overlay* (Song and Haithcoat, 2005) and *detection rate* (Lee et al., 2003), *correctness* C_{rp} and *quality* Q_{lp} from Rutzinger et al. (2009); *area omission error* A_{oe} and *area commission error* A_{ce} from Song and Haithcoat (2005) and *branching factor* B_f and *miss factor* M_f from Lee et al. (2003). A good building detection system should have high C_{mp} and C_{rp} values, but low A_{oe}, A_{ce}, B_f and M_f values, while $1 - C_{rp}$

indicates the false alarm rate of the system with respect to the building area. Q_{lp} makes a compromise between C_{mp} and C_{rp} (Heipke et al., 1997).

4.2.3. Geometric Index

As the geometry of the actual and detected buildings often differs significantly and the generally lower spatial resolution of the LIDAR data prohibits geometrically accurate building detection, the geometric evaluation system is rarely found in the literature. The shape similarity indices presented in Song and Haithcoat (2005) fall into this category and are application specific, for example, for cadastral management. Since it is assumed that both the reference and detected entities are rectangular, local changes in shapes are avoided and the shape indices are not considered.

Song and Haithcoat (2005) utilized root-mean-square-error (RMSE) values in order to estimate the geometric positional accuracy. For each *one-to-one* correspondence between detected and reference set, RMSE is measured as the average distance between a pair of detected and reference entities. Therefore, the RMSE is measured for TPs only, but not for FPs, FNs and MDs.

4.3. Data Sets

The test data set employed here was captured over Fairfield, NSW, Australia using an Optech laser scanner. Four sub-areas were used, the first covering an area of $248m \times 210m$ (Fig. 6(a)), the second covering an area of $155m \times 219m$ (Fig. 6(b)), the third covering an area of $228m \times 189m$ (Fig. 6(c)) and the fourth covering an area of $586m \times 415m$ (Fig. 6(d)). While the first two areas contain only residential buildings, the last two areas contain

both residential and industrial buildings. The first three areas were used for objective evaluation using 15 indices, while the fourth is a bigger area containing around 400 buildings and was used for visualization only. Last-pulse LIDAR data with a point spacing of $0.5m$ was used. A DEM (with $1m$ spacing) and four RGB colour orthophotos with a resolution of $0.15m$ were available for these areas. The fact that the orthoimage did not contain an infrared band was circumvented by computing a pseudo-NDVI image using the assumption that the three image bands are in the order of IR-Red-Green in order to be used in the standard NDVI formula (Kidwell, 1997).

The orthoimagery had been created using a bare-earth DEM, so that the roofs and the tree-tops were displaced with respect to the LIDAR data. Thus, data alignment was not perfect. Apart from this registration problem, there were also problems with shadows in the orthophotos, so the pseudo-NDVI image did not provide as much information as expected.

Reference data sets were created by monoscopic image measurement using the Barista software (Barista, 2009). All rectangular structures, recognizable as buildings and above the height threshold T_h (Section 4.1) were digitized. The reference data included garden sheds, garages, etc., that were sometimes as small as $10m^2$ in area. Altogether, 70, 62 and 60 buildings from the first three scenes formed the reference sets.

4.4. Results and Discussion

The algorithm was implemented and tested using Matlab 7.8.0 (R2009a) on a Windows XP machine with 3.00GHz of Intel(R) Core(TM)2 Duo CPU and 3.23GB of RAM. The average running time for first three scenes was about 13.5 minutes. The majority of time was taken up with accessing, load-

ing and storing the high volume of input and intermediate data. Memory-related limitations in Matlab precluded the possibility of conducting an experiment covering all 2400 buildings within the Fairfield data set. Such experimental validation will be possible once the algorithms are fully implemented within the Barista software.

The experimentation was carried out in two phases. Firstly, a sensitivity analysis of five important parameters (tile size, black pixel threshold, NDVI threshold, area reduction (reducing *length* and *width* of initial buildings) and colour similarity) was carried out to test how the detection algorithm performed when parameter values were changed. The standard parameter values were chosen for the test data sets. Secondly, the detection performance was evaluated using 15 indices in three categories when all the parameters were set at their chosen standard values.

4.4.1. Sensitivity Analysis

For sensitivity analysis five different values for each of the five parameters were used and object and pixel-based *qualities* were estimated. The reason for choosing *quality* as a measurement for sensitivity analysis is that it provides a balance between completeness and correctness (Heipke et al., 1997). The following values were used for the six parameters:

- *Tile size*: 400×400, 450×450, 500×500, 550×550 and 600×600 pixels;
- *Black pixel threshold*: 0.6, 0.7, 0.8, 0.9 and 1.0;
- *NDVI threshold*: 32, 40, 48, 56 and 64;

- *Area reduction*: 0.1, 0.15, 0.2, 0.25 and 3.0 and
- *Colour similarity*: 0.2, 0.3, 0.4, 0.5 and 0.6.

Fig. 8, in which the numbers 1 to 5 along the x-axis indicate the five values for each parameter, graphically illustrates the results. When one of the parameters was changed, others were set at their standard values. The pixel-based quality was given more weight than the object-based quality in the choice of the standard value for each parameter. Overall, one parameter - area reduction - was found to be moderately sensitive, while the other four were found less sensitive.

While the highest object-based quality was achieved at 20% area reduction and the highest pixel-based quality was achieved at 10% area reduction, at 15% area reduction both of these qualities were slightly lower than their highest values. Both the object- and pixel-based qualities were highest when the color similarity was 40%. For the tile size and NDVI threshold, the chosen values were 450×450 pixels and 48, respectively, when pixel-based qualities were highest and object-based qualities were slightly below the highest. An opposite scenario was observed when the black pixel threshold was 90%.

While a smaller tile size makes the mask generation procedure a bit expensive, a larger tile size may not clearly distinguish some buildings in a sloping tile because the estimated height threshold may not perfectly separate ground and above ground objects throughout the tile. A small NDVI threshold may remove some buildings as vegetation if building roofs have colours that are similar to trees. In contrast, a large NDVI threshold may detect some trees as buildings.

Table 1: Object-based evaluation results in percentages (C_m = completeness, C_r = correctness, Q_l = quality, M_d = multiple detection rate, D_o = Detection overlap rate, C_{rd} = detection cross-lap rate and C_{rr} = reference cross-lap rate).

Scenes	C_m	C_r	Q_l	M_d	D_o	C_{rd}	C_{rr}
Scene 1	97.14	97.14	95.31	2.60	6.85	2.74	7.14
Scene 2	95.94	96.55	92.08	4.62	5.00	1.67	4.84
Scene 3	98.33	99.25	90.94	11.60	28.57	14.29	33.33
Average	**97.14**	**97.9**	**92.78**	**6.27**	**13.47**	**6.23**	**15.11**

A smaller than 90% black pixel threshold may result in a nearby tree being included as a building part, for example. The same may happen if a more than 40% colour similarity is used when the building roof has a similar colour to the tree. While a smaller than 15% area reduction (*length* and *width* of initial buildings reduced by 15%) may not fully correct the registration error, a larger area reduction may stop the extension of the initial position if the roof has slightly different colours. This is why the pixel-based quality dropped more rapidly than the object-based quality, which indicates that though the buildings are correctly detected they are not correctly delineated. This is also evident from the evaluation results discussed below.

4.4.2. Evaluation using Standard Parameter Values

Table 1 shows the object-based evaluation results and Table 2 shows the pixel-based evaluation results. The geometric accuracy (RMSE) for the three scenes was $1.98m$, $1.91m$ and $1.86m$ with an average accuracy of 13 pixels ($1.92m$).

In object-based evaluation, more than 97% *completeness* and *correct-*

Table 2: Pixel-based evaluation results in percentages (C_{mp} = completeness, C_{rp} = correctness, Q_{lp} = quality, A_{oe} = area omission error, A_{ce} = area commission error, B_f = branching factor and M_f = miss factor).

Scenes	C_{mp}	C_{rp}	Q_{lp}	A_{oe}	A_{ce}	B_f	M_f
Scene 1	77.32	89.29	70.07	22.68	10.35	12.00	29.33
Scene 2	77.97	87.05	67.40	22.03	12.67	14.87	28.26
Scene 3	79.51	90.35	72.11	20.49	7.54	10.68	25.77
Average	**78.27**	**88.90**	**69.86**	**21.74**	**10.19**	**12.52**	**27.79**

ness resulted in an average 92% *quality* with at least 6% of buildings being detected multiple times. The *reference cross-lap rate* was higher than the *detection cross-lap rate*, since some nearby trees were detected along with the actual buildings. In pixel-based evaluation, while 78% of building areas were completely detected, resulting in a 21% *omission error*, 89% of detected areas were correct, offering a 10% *commission error*. Since the *miss factor* and *omission error* were larger than the *branching factor* and *commission error*, respectively, the *false positive rate* of the proposed technique is lower than its *false negative rate*.

Overall, in both object- and pixel-based evaluations, the proposed detection technique performed better on Scene 1 than on Scene 2 in terms of all indices except *cross-lap* and *detection overlap* rates. There were two reasons for this: a) some true buildings were detected twice in Scene 1, and b) in Scene 1 though all true buildings were detected with some of them being missed partially, some false buildings (actually trees) were also detected. Scene 3 performed better than Scenes 1 and 2 in pixel-based evaluation whereas Scene

3 gave higher *cross-lap* and *detection overlap rates* in object-based evaluation due to multiple detection of complex industrial buildings. In the geometric evaluation, in terms of RMSE, there was at least $0.05m$ better positional accuracy for Scene 3 than for Scenes 1 and 2.

It was found that the use of NDVI (actually pseudo NDVI in this case) did not perform well in distinguishing between trees and building roofs, especially when both were of similar colour. While a low NDVI threshold removed some true buildings, a high NDVI threshold detected some trees as buildings. The difference in first- and last-pulse LIDAR data was also investigated to remove trees, as was done in Rottensteiner et al. (2005), but was found less useful. The outcome regarding NDVI and the difference in first and last pulse data supports the finding of an earlier study by Rottensteiner et al. (2007).

Since different published detection techniques follow different evaluation systems on different data sets, they are difficult to compare. As with the proposed detection technique, Sohn and Dowman (2007) and Cheng et al. (2008) also used line segments and building geometry adjustment using dominant line angles. Unlike Cheng et al. (2008) and the proposed technique, Sohn and Dowman (2007) used specific building models to fit the LIDAR points. While the proposed technique introduces a threshold-free evaluation system, both of these existing techniques employ threshold-based evaluation systems.

In terms of object-based *correctness* the proposed technique performed much better than that of Cheng et al. (2008). The method of Sohn and Dowman (2007) offered slightly higher pixel-based performance than the proposed technique because of the adopted evaluation system (Rottensteiner et al.,

2005), which excluded FP and FN buildings from evaluation and established many-to-many relationships between the detected and reference sets. Establishing one-to-one correspondences by the proposed detector increases the number of FP and FN buildings and the proposed evaluation system considers all of them.

The same Fairfield data set was previously employed by Rottensteiner et al. (2005), Rottensteiner et al. (2007) and Rutzinger et al. (2009) to investigate automated building extraction. However, in those investigations, two different threshold-based evaluation systems were employed and the Dempster-Shafer (DS) detector was evaluated using *completeness*, *correctness* and *quality*. Rutzinger et al. (2009) has presented results of pixel-based evaluation of the DS detector showing that it can offer higher *completeness* (92.1%) and *quality* (81.8%) than the proposed detector. However, in object-based evaluation the DS detector offered much lower *completeness* (44.2%) and *quality* (43.1%) than the proposed detector. The superior performance of the DS detector in pixel-based evaluation was largely due to the adopted evaluation systems, Rottensteiner et al. (2005) and Rutzinger et al. (2009)) which excluded FP and FN buildings from evaluation and established many-to-many relationships between the detected and reference sets. Moreover, unlike the proposed detector the DS detector was excessively sensitive to small buildings (performance deteriorated with the decrease of building size) and buildings smaller than $30m^2$ could not be detected (Rottensteiner et al., 2007).

5. Conclusion

This paper has proposed an automatic building detection technique using LIDAR data and multispectral imagery. The initial building positions are obtained from the primary building mask derived from LIDAR data. The final building positions are obtained by extending their initial positions based on colour information, and the two masks ensure the accurate delineation of the buildings. In particular, the primary building mask helps separate building detections when they are very close to each other and the secondary building mask helps to confine the extension of initial positions outside a building when the roof and ground have similar colour information.

Experimental testing has shown that the proposed technique can detect urban residential and industrial buildings of different shapes with a very high success rate. However, the technique can display shortcomings in areas of high-terrain slope and those with dense high-rise buildings of rapidly varying height within a given tile size, since in such areas the average DEM height may not necessarily correspond to the actual ground height. Extension of the algorithm's functionality to better accommodate such situations is currently under investigation.

Another important observation from the presented results is that object-based *completeness* (*detection rate* 97%) is high when compared to pixel-based *completeness* (*matching overlay* 78%). However, the geometric positional accuracy remains relatively poor (13 pixels) for mapping purposes; although not for applications where building detection is the primary goal. This observation indicates that some of the truly detected buildings are not completely delineated due to small local variations along the roof boundary,

occlusion by nearby trees or different roof colours in and out of the initial building position. Consequently, the proposed detection technique can be applied in city planning, homeland security, disaster (flood or bushfire) management and building change detection with high reliability, but it is not as yet applicable to cadastral mapping and accurate roof plane extraction, both of which require higher pixel-based and geometric accuracy.

Acknowledgment

The authors would like to thank AAMHatch (www.aamhatch.com.au) for providing the Fairfield data set. Special thanks are due to Dr Franz Rottensteiner of the Institute of Photogrammetry and GeoInformation, Leibniz University Hannover for his insightful advice.

References

Awrangjeb, M., Lu, G., Fraser, C. S., Ravanbakhsh, M., 2009. A fast corner detector based on the chord-to-point distance accumulation technique. In: Proc. Digital Image Computing: Techniques and Applications. Melbourne, Australia, pp. 519–525.

Barista, 2009. The barista software.
`www.baristasoftware.com.au`

Canny, J., 1986. A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 8 (6), 679–698.

Chen, L., Teo, T., Shao, Y., Lai, Y., Rau, J., 2004. Fusion of LIDAR data and optical imagery for building modelling. International Archives of Pho-

767 togrammetry, Remote Sensing and Spatial Information Sciences 35 (part
768 B4), 732–737.

769 Cheng, L., Gong, J., Chen, X., Han, P., 2008. Building boundary extraction
770 from high resolution imagery and LIDAR data. International Archives of
771 the Photogrammetry, Remote Sensing and Spatial Information Sciences
772 37 (part B3), 693–698.

773 Demir, N., Poli, D., Baltsavias, E., 2009. Extraction of buildings using im-
774 ages & LIDAR data and a combination of various methods. International
775 Archives of the Photogrammetry, Remote Sensing and Spatial Information
776 Sciences 38 (part 3/W4), 71–76.

777 Foody, G., 2002. Status of land cover classification accuracy assessment. Re-
778 mote Sensing of Environment 80 (1), 185–201.

779 Haala, N., Brenner, C., 1999. Extraction of buildings and trees in urban
780 environments. ISPRS Journal of Photogrammetry and Remote Sensing
781 54 (2-3), 130–137.

782 Heipke, C., Mayer, H., Wiedemann, C., Jamet, O., 1997. Evaluation of au-
783 tomatic road extraction. International Archives of Photogrammetry and
784 Remote Sensing 32 (part 3-2W3), 47–56.

785 Khoshelham, K., Nedkov, S., Nardinocchi, C., 2008. A comparison of
786 bayesian and evidence-based fusion methods for automated building detec-
787 tion in aerial data. International Archives of the Photogrammetry, Remote
788 Sensing and Spatial Information Sciences 37 (part B7), 1183–1188.

Kidwell, K., 1997. NOAA Global Vegetation Index User's Guide. U.S. Department of Commerce NOAA National Environmental Satellite data and Information Service.

Lee, D., Lee, K., Lee, S., 2008. Fusion of LIDAR and imagery for reliable building extraction. Photogrammetric Engineering and Remote Sensing 74 (2), 215–226.

Lee, D., Shan, J., Bethel, J., 2003. Class-guided building extraction from ikonos imagery. Photogrammetric Engineering and Remote Sensing 69 (2), 143–150.

Mayer, H., 1999. Automatic object extraction from aerial imagery - a survey focusing on buildings. Computer Vision and Image Understanding 74 (2), 138–149.

Oude Elberink, S., 2008. Problems in automated building reconstruction based on dense airborne laser scanning data. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 37 (part B3), 93–98.

Rottensteiner, F., Trinder, J., Clode, S., Kubik, K., 2005. Using the Dempster Shafer method for the fusion of LIDAR data and multi-spectral images for building detection. Information Fusion 6 (4), 283–300.

Rottensteiner, F., Trinder, J., Clode, S., Kubik, K., 2007. Building detection by fusion of airborne laser scanner data and multi-spectral images : Performance evaluation and sensitivity analysis. ISPRS Journal of Photogrammetry and Remote Sensing 62 (2), 135–149.

812 Rutzinger, M., Rottensteiner, F., Pfeifer, N., 2009. A comparison of evalua-
813 tion techniques for building extraction from airborne laser scanning. IEEE
814 Journal of Selected Topics in Applied Earth Observations and Remote
815 Sensing 2 (1), 11–20.

816 Sampath, A., Shan, J., 2007. Building boundary tracing and regularization
817 from airborne LIDAR point clouds. Photogrammetric Engineering and Re-
818 mote Sensing 73 (7), 805–812.

819 Shan, J., Lee, S., 2005. Quality of building extraction from IKONOS imagery.
820 ASCE Journal of Surveying Engineering 131 (1), 27–32.

821 Shufelt, J., 1999. Performance evaluation and analysis of monocular building
822 extraction from aerial imagery. IEEE Transactions on Pattern Analysis
823 and Machine Intelligence 21 (4), 311–326.

824 Sohn, G., Dowman, I., 2007. Data fusion of high-resolution satellite imagery
825 and LIDAR data for automatic building extraction. ISPRS Journal of Pho-
826 togrammetry and Remote Sensing 62 (1), 43–63.

827 Song, W., Haithcoat, T., 2005. Development of comprehensive accuracy as-
828 sessment indexes for building footprint extraction. IEEE Transactions on
829 Geoscience and Remote Sensing 43 (2), 402–404.

830 Sun, J., Lin, Y., Kang, S., Shum, H., 2005. Symmetric stereo matching for
831 occlusion handling. In: Proc. IEEE Conference on Computer Vision and
832 Pattern Recognition. Vol. 2. San Diego, CA, USA, pp. 399–406.

833 Vu, T., Yamazaki, F., Matsuoka, M., 2009. Multi-scale solution for building

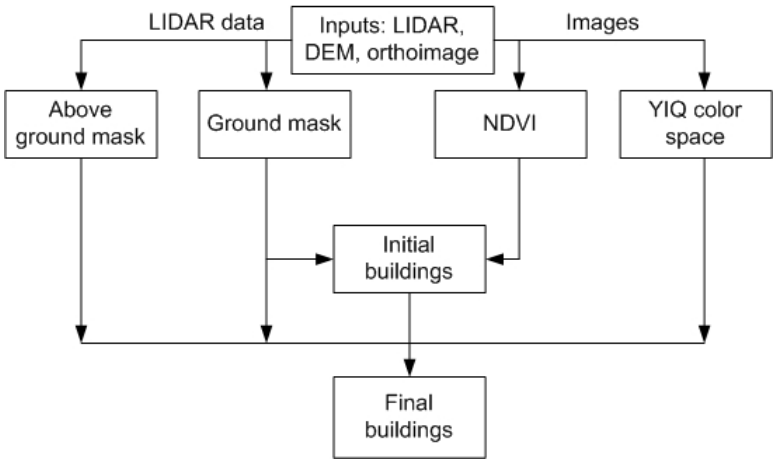

Figure 1: Flow diagram of the proposed building detection technique.

834 extraction from LIDAR and image data. International Journal of Applied
835 Earth Observation and Geoinformation 11 (4), 281–289.

836 Yong, L., Huayi, W., 2008. Adaptive building edge detection by combining
837 LIDAR data and aerial images. International Archives of the Photogram-
838 metry, Remote Sensing and Spatial Information Sciences 37 (part B1),
839 197–202.

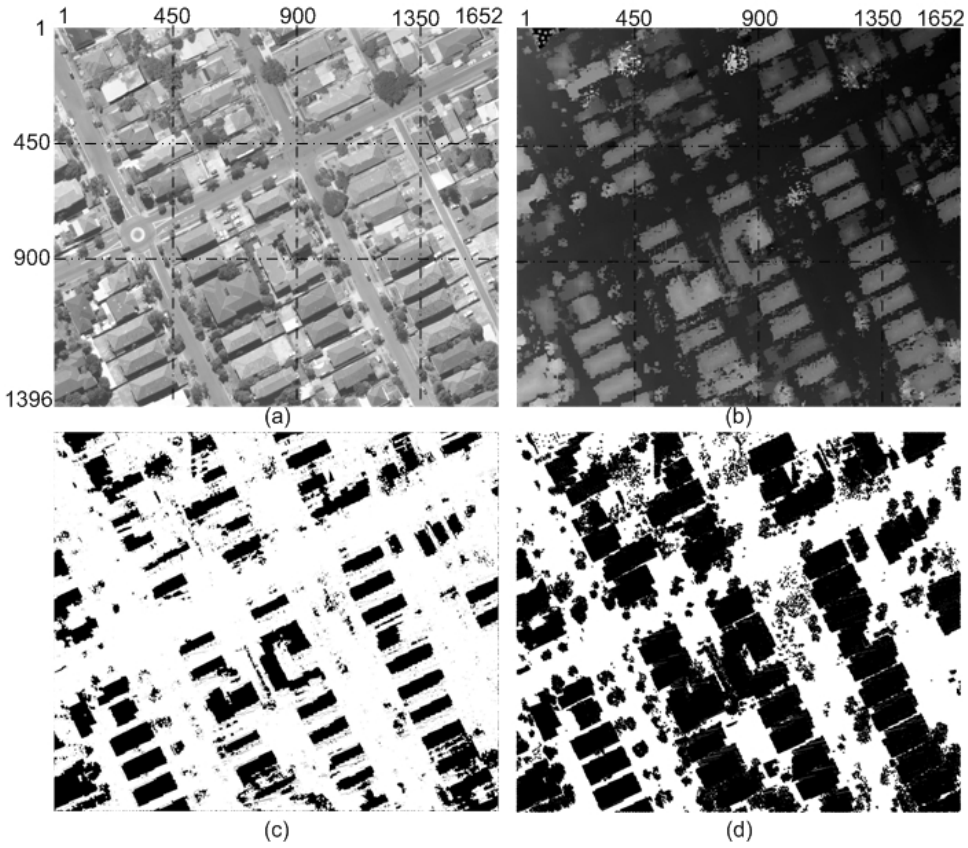

Figure 2: (a) A test scene, (b) LIDAR data (shown in gray-scale), (c) primary building mask and (d) secondary building mask.

Figure 3: (a) Application of NDVI and (b) line-adjustment to remove tree-edges. Lines with small circles at centres are removed. (c) Initial building positions.

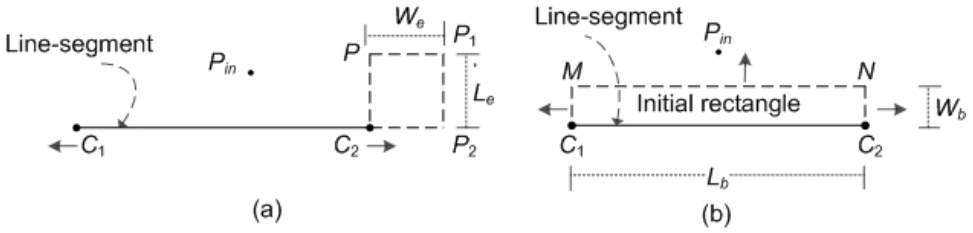

Figure 4: (a) Extending a line segment (b) Forming an initial building by extending three sides of a rectangle on a line segment. Arrows indicate extension directions.

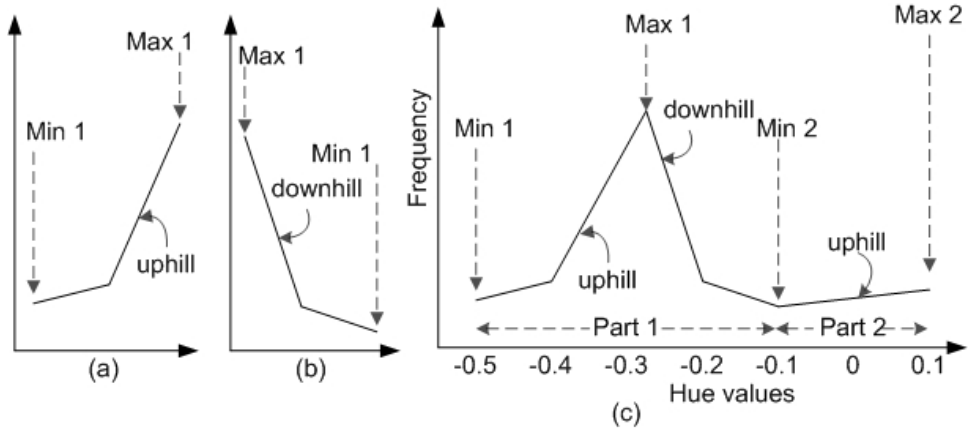

Figure 5: Basic histograms: (a) uphill and (b) downhill. (c) a combination of basics uphill-downhill-uphill.

Figure 6: Detected buildings on the orthoimages.

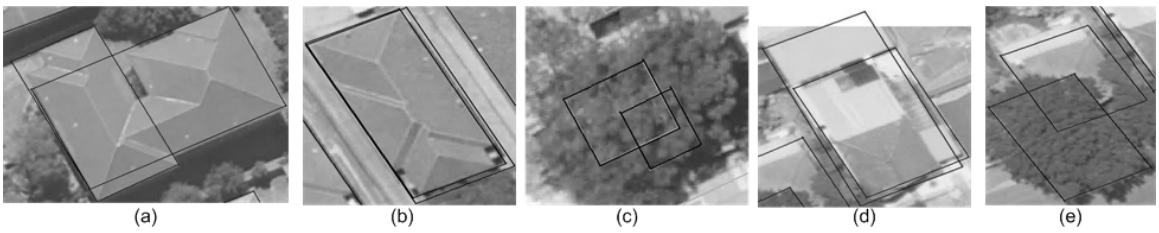

Figure 7: Different types of detection overlaps: (a) natural, (b) multiple detection, (c) false-false, (d) true-true and (e) true-false.

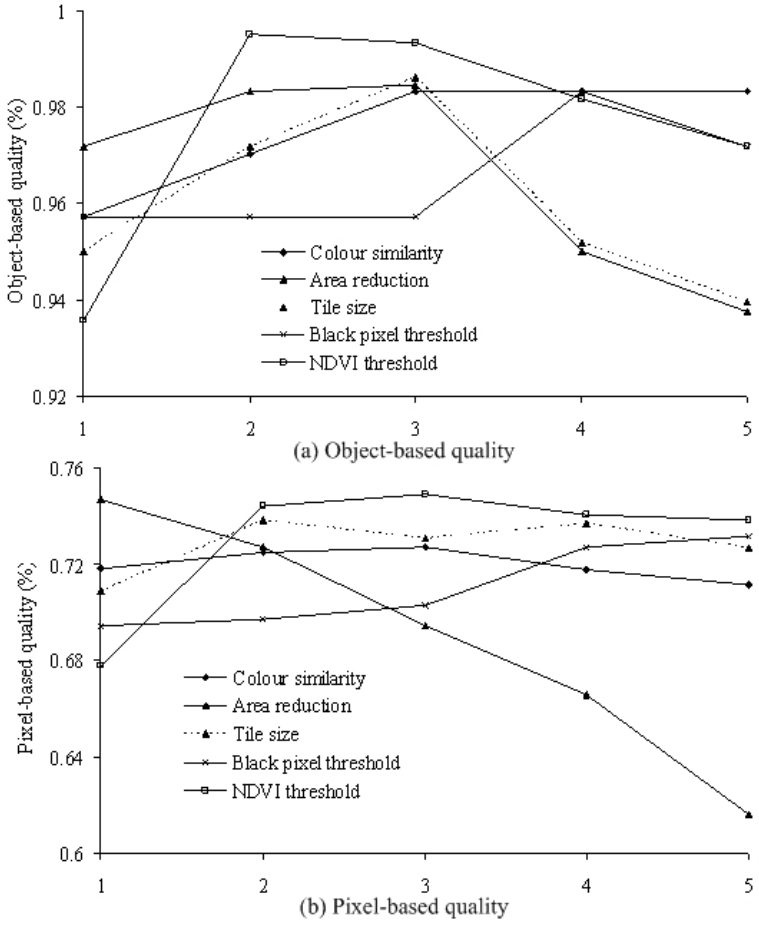

Figure 8: Sensitivity of different parameters.