

**Automatic building detection using LIDAR data and multispectral imagery**

**Author**

Awrangjeb, M, Ravanbakhsh, M, Fraser, CS

**Published**

2010

**Conference Title**

Digital Image Computing: Techniques and Applications (DICTA), Proceedings

**Version**

Accepted Manuscript (AM)

**DOI**

[10.1109/DICTA.2010.17](https://doi.org/10.1109/DICTA.2010.17)

**Downloaded from**

<http://hdl.handle.net/10072/392087>

**Griffith Research Online**

<https://research-repository.griffith.edu.au>

# Automatic Building Detection Using LIDAR Data and Multispectral Imagery

Mohammad Awrangjeb, Mehdi Ravanbakhsh and Clive S. Fraser

Department of Geomatics, The University of Melbourne  
CRC for Spatial Information, 723 Swanston St, Carlton Vic 3053 Australia

{mawr, m.ravanbakhsh, c.fraser}@unimelb.edu.au

## Abstract

*An automatic building detection technique using LIDAR data and multispectral imagery has been proposed. 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 buildings, when assessed in terms of 15 indices including completeness, correctness and quality.*

## 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 [5].

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 [17]. Firstly, 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 belongs to non-building objects. Secondly, there is always a significant loss of relevant building cues due to occlusions, poor contrast, shadows and disadvantageous image perspec-

tive. Finally, the primary data to support building detection is available from a variety of sources with different resolution, each source having its own advantages and disadvantages.

Lee et al. [11] have categorized building detection techniques into three groups. The first group of algorithms uses 2D or 3D information from photogrammetric imagery [13]. These algorithms are complex due to involvement of detailed information in high-resolution images [2] and complicated and erroneous estimation of 3D (height) information [19]. Algorithms in the second group consider building detection as a classification problem and detect building regions from LIDAR (LIght Detection And Ranging) data [11]. However, the use of raw or interpolated data can influence the detection performance [6] resulting in poor horizontal accuracy for building edges [21]. As a result, it can be hard to obtain a detailed and geometrically precise boundary using only LIDAR point clouds [2].

The introduction of LIDAR has, however, offered a favourable option for improving the level of automation in the building detection process when compared to image-based detection [20]. The third category of methods uses both LIDAR data and photogrammetric imagery, since each have unique attributes for building detection and the advantages of one can compensate for disadvantages of the other. More specifically, intensity and height information in LIDAR data can be used with texture and region boundary information in aerial imagery to improve accuracy [11]. However, 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 [21]; only a few approaches with technical details have thus far been published [14].

This paper aims at a successful integration of LIDAR data and photogrammetric imagery for building detection so that improved detection performance is obtained. Two masks are generated from the LIDAR data – primary and secondary building masks. The initial buildings are formed around the black regions in the primary mask. The possibly false buildings around trees are avoided using the normal-

ized difference vegetation index (NDVI) image. The final buildings are obtained by extending their initial positions using the secondary mask and colour information of the multispectral imagery. The performance of the proposed building detection approach has been assessed using an automatic threshold-free evaluation system involving 15 indices.

This paper, which is a condensed version of [2] with additional experimental validations, is organized as follows. Section 2 presents a review of the integration techniques of LIDAR data and imagery. Section 3 presents the proposed building detection technique. Section 4 presents the performance evaluation and, finally, Section 5 offers concluding remarks.

## 2. Related Work

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 [14, 20]. As a result, they can suffer from poor horizontal accuracy for the detected buildings. Rotensteiner et al. [14] 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 [15]. The reason is that if the Dempster-Shafer model is not properly trained, then the misclassification rate increases considerably [9]. Vu et al. [20] 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 [8, 4, 17, 11, 6] which use both LIDAR data and imagery as primary cues to delineate building outlines. They also employ imagery to remove vegetation. Consequently, they offer better horizontal accuracy for detected buildings. The proposed building detection technique falls into this group. Haala and Brenner [8] 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. [4] 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 [17] and could not detect small buildings.

Sohn and Dowman [17] 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. [5] proposed a similar technique with precise geometric position. Lee et al.

[11] 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. [6] 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.

## 3. Proposed Detection Technique

Fig. 1 shows the flow diagram of the proposed building detection technique. The input information consists of a LIDAR point cloud and multispectral orthoimagery. The primary and secondary 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 which is usually in the RGB system, 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.

### 3.1. Generation of Masks

The raw LIDAR data is divided into groups, where each group corresponds to a tile of  $450 \times 450$  image pixels; i.e., all laser points corresponding to an image tile go into the same group. A histogram of the height data for each LIDAR group is obtained, where bins of low heights correspond to ground areas and high points correspond to trees and buildings. The distance between successive bins is  $2m$  and the bin having the maximum frequency indicates the ground height  $H_g$  for the corresponding tile. This is based on the assumption that the majority of the LIDAR points have heights similar to the ground height. Alternatively, the average DEM value in each tile can be used as  $H_g$ . Figs. 2(a)-(b) show the tiles of masks on an orthoimage and the groups of LIDAR data.

A threshold  $T_h = H_g + 2.5m$  is applied to split the LIDAR points into two sets: lower and higher point heights. The first set marks white for each of its points in the primary building mask  $M_p$ , which is initially a completely black mask. The second set marks black for each of its points in the secondary building mask  $M_s$ , which is initially a completely white mask. Consequently, the black areas in the primary building mask indicate *void areas* where there are no laser returns below  $T_h$  and those in the secondary building mask indicate *filled areas* from where returns indicate an elevated object above the same height threshold. Figs. 2(c)-(d) show the two generated masks for a test scene.

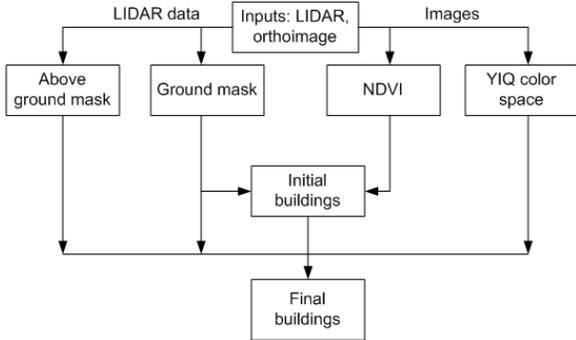


Figure 1. Flow diagram of the proposed building detection technique.

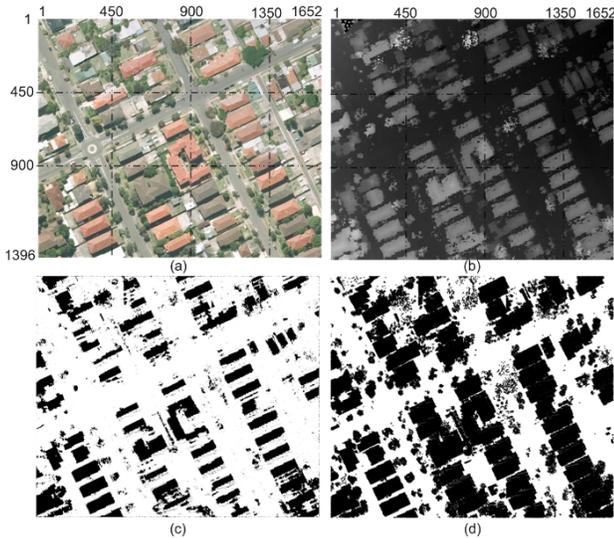


Figure 2. (a) A test scene, (b) LIDAR data (shown in gray-scale), (c) primary building mask and (d) secondary building mask.

### 3.2. Initial Buildings

Initial buildings are the black areas in the primary building mask as shown in Fig. 2(c). Three steps are followed to obtain these black regions. Firstly, lines around the black shapes in  $M_p$  are formed. Secondly, the lines are adjusted and extended. Finally, rectangular shapes are obtained using these lines.

Edges are first extracted from  $M_p$  using an edge detector and short edges are removed assuming that the minimum building length or width is  $3m$ . Corners are then detected on each curve using the fast corner detector in [1]. 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. In order to properly 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.

In order to avoid detected tree-edges, the mean of the NDVI value  $\Upsilon$  is calculated on both sides of each line segment. For a 4-band (RGBI) image the NDVI image is calculated using the standard formula:

$$NDVI = \frac{I - R}{I + R}, \quad (1)$$

where  $I$  indicates near infrared and  $R$  the red band. If  $\Upsilon$  is above a threshold  $T_{ndvi} = 10$  for any side, the line segment is classed as a tree-edge and removed. In the case of a 3-band (RGB) image, since there is no infrared band, the pseudo-NDVI image is calculated using the assumption that the 3 bands are in the order of I-R-G such that they can be used in the above standard NDVI formula [10]. In fact, we estimate  $\Upsilon$  from the sigma (variance) of the pseudo-NDVI image for a 3-band image. If  $\Upsilon$  is above a threshold  $T_{ndvi} = 48$  for any side, the line segment is classed as a tree-edge and removed.

In the second step, the line segments are adjusted and extended. The adjustment is based on the assumption that longer lines are more likely to be building edges. In an iterative procedure starting from the longest line and taking it as a reference, the angle between the reference and each line in its neighbourhood is estimated. The lowest rotation angle  $\theta_r$  is recorded for each line over all iterations (for all long lines taken as references). After the iterative procedure, each line is rotated with respect to its centre by  $\theta_r$ . If a rotation angle is not recorded for a line it is removed as a tree-edge. Each adjusted line is then extended iteratively by considering the percentage of black pixels (more than 70%) and applying the NDVI threshold to the building side.

Finally, initial buildings are formed among the extended line segments. 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. Before detecting a rectangle using a line segment in each iteration, the line segment is first tested to ascertain whether it is already in a detected building. In order to detect an initial building on a line segment, an initial rectangle (of width  $1.5m$ ) is formed on the building side and then three of its sides are extended outwards with respect to  $P_{in}$  using the same technique applied to extend the extracted lines. Fig. 3(a) shows the initial detected buildings on the test scene.

### 3.3. Final Buildings

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



Figure 3. (a) Initial and (b) Final buildings.

than one actual buildings, and  $M_s$  is used to avoid unexpected extension of an initial position beyond the actual building roof.

An initial building position may extend beyond 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 are estimated using colour histograms 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.

Moreover, since black areas in  $M_p$  represents the void areas where the laser does not reach below a certain height threshold, an initial building position may be detected on a swimming pool or a river. In order to avoid such false buildings, the percentage of black pixels within each initial building position is calculated from  $M_s$  and if it is less than 70%, the initial building is rejected.

There may be overlaps between the detected initial positions. 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.

The initial building positions are sorted in descending order of their *length* or *area*, since both of these sorted lists were found to offer the same performance. Starting from the initial building having the longest *length* or largest *area*, its four sides are extended outwards separately. While extending each side in an iterative procedure, the percentages of black pixels in both  $M_p$  and  $M_s$  and of dominant colour components within the estimated colour threshold pairs are estimated. The side is extended if percentages of black pixels are above 90% and those of dominant colour components are above 40%. Fig. 3(b) shows the final detected buildings on the test scene.

## 4. Performance Evaluation

The adopted 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. Altogether 15 indices are used in three categories (object-based, pixel-based and geometric) to evaluate the performance. Most of the indices have been adopted from the literature and the rest are proposed for a more complete evaluation.

### 4.1. Evaluation System

Two sets of data, 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 ‘T’ shapes, were used for evaluation. These two sets are the *detected set*, which is obtained from the proposed automatic building detection technique, and the *reference set*, which is obtained from manual building measurement within the orthoimagery. In an approach similar to that of [18], a detected entity is counted as correct if any of its part overlaps a reference entity. However, unlike existing evaluation systems [14, 16], 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

Seven indices are used for object-based evaluation. *Completeness*  $C_m$ , also known as *detection rate* [18] or *producer’s accuracy* [7]; *correctness*  $C_r$ , also known as *user’s accuracy* [7]; and *quality*  $Q_l$  have been adopted from [16]. *Multiple detection rate* is the percentage of multiply and correctly detected entities in the detected set. *Detection overlap rate* is the percentage of overlap in the detected set. *Detection cross-lap rate* is defined as the percentage of detected entities which overlap more than one reference entity. *Reference cross-lap rate* is defined as the percentage of reference entities which are overlapped by more than one detected entity (see [2] for formal definitions).

A total of 7 pixel-based evaluation indices are also used, these being: *completeness*  $C_{mp}$ , also known as *matched*

overlay [18] and detection rate [12]; correctness  $C_{rp}$  and quality  $Q_{lp}$  from [16]; area omission error  $A_{oe}$  and area commission error  $A_{ce}$  from [18]; and branching factor  $B_f$  and miss factor  $M_f$  from [12].

Root-mean-square-error (RMSE) values [18] indicate the geometric positional accuracy. For each one-to-one correspondence between detected and reference set, the RMSE is measured as the average distance between a pair of detected and reference entities. Therefore, the RMSE is measured for true positive buildings only.

### 4.3. Data Sets

The test data sets employed here were cover Fairfield, NSW and Moonee Ponds, Victoria. While the Fairfield data set covers an area of  $588m \times 417m$  containing 370 buildings, the Moonee Ponds data set covers an area of  $400m \times 400m$  containing 250 buildings. For Fairfield, last-pulse LIDAR data with a point spacing of  $0.5m$ , a DEM (with  $1m$  spacing) and an RGB colour orthophoto with a resolution of  $0.15m$  were available. For Moonee Ponds, first-pulse LIDAR data with a point spacing of  $1m$ , a DEM (with  $1m$  spacing) and an RGBI colour orthophoto with a resolution of  $0.1m$  were available. In fact, for Moonee Ponds the whole RGBI image consisted of four sub-images (samples), each comprising  $2000 \times 2000$  pixels.

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 NDVI and pseudo-NDVI images did not provide as much information as expected.

Reference data sets were created by monoscopic image measurement using the Barista software [3]. All rectangular structures, recognizable as buildings and above the height threshold  $T_h$  were digitized. The reference data included garden sheds, garages, etc. These were sometimes as small as  $10m^2$  in area.

### 4.4. Results and Discussion

Figs. 4 and 5 show building detection results on the whole Fairfield area and a sample of Moonee Ponds. For each area, the detection results are shown when the ground thresholds were estimated using histogram maxima and DEM average (see Section 3.1). In both of the cases, the proposed building detection algorithm offered almost the same visual performance.

Table 1 shows the object-based, Table 2 the pixel-based and Table 3 the geometric evaluation results. In the object-based evaluation, more than 93% completeness and correctness resulted in an average 89% quality with at least 5% of buildings being detected multiple times. The reference cross-lap rate was higher than the detection cross-lap rate,



Figure 4. Building detection in Fairfield sample. Ground thresholds were estimated using: (a) histogram maxima and (b) average DEM.

Table 1. Object-based evaluation results in percentages (FF = Fairfield, MP = Moonee Ponds,  $C_m$  = 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	$C_m$	$C_r$	$Q_l$	$M_d$	$D_o$	$C_{rd}$	$C_{rr}$
FF, Hist.	95.1	95.1	90.7	6.1	18.3	17.5	28.7
FF, DEM	92.2	90.8	88.8	3.9	14.4	21.6	27.8
MP, Hist.	96.6	93.5	90.6	6.3	13.8	9.9	15.3
MP, DEM	92.6	94.5	87.8	7.0	14.2	8.7	18.1
<b>Average</b>	<b>94.1</b>	<b>93.5</b>	<b>89.5</b>	<b>5.8</b>	<b>15.2</b>	<b>14.4</b>	<b>22.5</b>

since some nearby trees were detected along with the actual buildings. In the pixel-based evaluation, while 83% of building areas were completely detected, resulting in a 17% omission error, 88% of detected areas were correct, offering at least a 14% 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



(a)



(b)

Figure 5. Building detection in Moonee Ponds sample. Ground thresholds were estimated using: (a) histogram maxima and (b) average DEM.

proposed technique is lower than its *false negative rate*.

Overall, in object-based, pixel-based and geometric evaluations, the histogram-based ground thresholds offered better performance than the DEM-based ground thresholds. This indicates that the classification results and the interpolation technique used in the calculation of the DEM grid may not represent the actual ground height for each tile. In the object-based evaluation, the detection results

Table 2. Pixel-based evaluation results in percentages (FF = Fairfield, MP = Moonee Ponds,  $C_{mp}$  = 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	$C_{mp}$	$C_{rp}$	$Q_{lp}$	$A_{oe}$	$A_{ce}$	$B_f$	$M_f$
FF, Hist.	81.4	85.1	71.3	18.6	14.1	17.5	22.9
FF, DEM	78.2	81.0	66.1	21.8	15.5	23.4	27.9
MP, Hist.	86.9	85.6	75.7	13.1	13.5	17.0	15.1
MP, DEM	85.2	83.6	73.1	14.9	14.8	18.0	19.5
<b>Average</b>	<b>82.9</b>	<b>83.8</b>	<b>71.6</b>	<b>17.1</b>	<b>14.5</b>	<b>19.0</b>	<b>21.4</b>

Table 3. Geometric accuracy (FF = Fairfield, MP = Moonee Ponds).

Scenes	in metres	in pixels
FF, Hist.	2.4	16.0
FF, DEM	2.8	18.7
MP, Hist.	1.7	17.0
MP, DEM	1.8	18.0
<b>Average</b>	<b>2.2</b>	<b>17.4</b>

were almost the same for both data sets, but for pixel-based and geometric evaluations the proposed detector performed slightly better for Moonee Ponds than for Fairfield. This was because of availability of 4-band higher resolution imagery for the Moonee Ponds data set. In other words, the NDVI image (calculated for 4-band Moonee Ponds orthophotos) has higher discriminating power than the pseudo-NDVI image (calculated for 3-band Fairfield orthophotos) and higher resolution imagery offers better colour separation of roof boundaries.

The same Fairfield data set was previously employed by [14], [15] and [16] 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. [16] has presented results of pixel-based evaluation of the DS detector showing that it can offer higher *completeness* and *quality* than the proposed detector. However, in object-based evaluation the DS detector offered much lower *completeness* and *quality* than the proposed detector. The superior performance of the DS detector in pixel-based evaluation was largely due to the adopted evaluation systems ([14] and [16]), which excluded false-positive and false-negative 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 a decrease of building size) and buildings smaller than  $30m^2$  could not be detected [15].

## 5. Conclusion

An automatic building detection technique using LIDAR data and multispectral imagery has been described. 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 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 rectilinear buildings of different shapes with a very high success rate.

An important observation from the presented results is that object-based *completeness* (detection rate 94.1%) is high when compared to pixel-based *completeness* (matching overlay 82.9%). However, the geometric positional accuracy remains relatively poor (17.4 pixels) for mapping purposes; although not for applications where building detection is the primary goal. This observation suggests that the proposed detection technique can be applied with high reliability in city planning, homeland security, disaster (flood or bushfire) management and building change detection, but it is not as yet applicable to cadastral mapping and accurate roof plane extraction, both of which require higher geometric accuracy.

## References

- [1] M. Awrangjeb, G. Lu, C. S. Fraser, and M. Ravanbakhsh. A fast corner detector based on the chord-to-point distance accumulation technique. In *Proc. Digital Image Computing: Techniques and Applications*, pages 519–525, Melbourne, Australia, Dec 2009.
- [2] M. Awrangjeb, M. Ravanbakhsh, and C. S. Fraser. Automatic detection of residential buildings using lidar data and multispectral imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2010.
- [3] BaristaSoftware. [www.baristasoftware.com.au](http://www.baristasoftware.com.au), 2010.
- [4] L. Chen, T. Teo, Y. Shao, Y. Lai, and J. Rau. Fusion of lidar data and optical imagery for building modelling. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 35(part B4):732–737, 2004.
- [5] L. Cheng, J. Gong, X. Chen, and P. Han. Building boundary extraction from high resolution imagery and lidar data. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37(part B3):693–698, 2008.
- [6] N. Demir, D. Poli, and E. Baltsavias. Extraction of buildings using images & lidar data and a combination of various methods. *Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 38(part 3/W4):71–76, 2009.
- [7] G. Foody. Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80(1):185–201, 2002.
- [8] N. Haala and C. Brenner. Extraction of buildings and trees in urban environments. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(2-3):130–137, 1999.
- [9] K. Khoshelham, S. Nedkov, and C. Nardinocchi. A comparison of bayesian and evidence-based fusion methods for automated building detection in aerial data. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37(part B7):1183–1188, 2008.
- [10] K. Kidwell. *Global Vegetation Index User's Guide*. NOAA/NESDIS, U.S. Department of Commerce, Washington, D.C., 1997.
- [11] D. Lee, K. Lee, and S. Lee. Fusion of lidar and imagery for reliable building extraction. *Photogrammetric Engineering and Remote Sensing*, 74(2):215–226, 2008.
- [12] D. Lee, J. Shan, and J. Bethel. Class-guided building extraction from ikonos imagery. *Photogrammetric Engineering and Remote Sensing*, 69(2):143–150, 2003.
- [13] H. Mayer. Automatic object extraction from aerial imagery - a survey focusing on buildings. *Computer Vision and Image Understanding*, 74(2):138–149, 1999.
- [14] F. Rottensteiner, J. Trinder, S. Clode, and K. Kubik. Using the dempstershafer method for the fusion of lidar data and multi-spectral images for building detection. *Information Fusion*, 6(4):283–300, 2005.
- [15] F. Rottensteiner, J. Trinder, S. Clode, and K. Kubik. 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, 2007.
- [16] M. Rutzinger, F. Rottensteiner, and N. Pfeifer. A comparison of evaluation techniques for building extraction from airborne laser scanning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2(1):11–20, 2009.
- [17] G. Sohn and I. Dowman. Data fusion of high-resolution satellite imagery and lidar data for automatic building extraction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(1):43–63, 2007.
- [18] W. Song and T. Haithcoat. Development of comprehensive accuracy assessment indexes for building footprint extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 43(2):402–404, 2005.
- [19] J. Sun, Y. Lin, S. Kang, and H. Shum. Symmetric stereo matching for occlusion handling. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 399–406, San Diego, CA, USA, 2005.
- [20] T. Vu, F. Yamazaki, and M. Matsuoka. Multi-scale solution for building extraction from lidar and image data. *International Journal of Applied Earth Observation and Geoinformation*, 11(4):281–289, 2009.
- [21] L. Yong and W. Huayi. Adaptive building edge detection by combining lidar data and aerial images. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37(part B1):197–202, 2008.