BUILDING DETECTION FROM MULTISPECTRAL IMAGERY AND LIDAR DATA EMPLOYING A THRESHOLD-FREE EVALUATION SYSTEM

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
E-mails: {mawr, m.ravanbakhsh, c.fraser}@unimelb.edu.au
Commission III-WG III/4

KEY WORDS: Building detection, LIDAR, point cloud, multispectral, photogrammetry imagery, orthoimage, fusion, edge feature

ABSTRACT:

This paper presents an automatic system for the detection of buildings from LIDAR data and multispectral imagery, which employs a threshold-free evaluation system that does not involve any thresholds based on human choice. Two binary masks are obtained from the LIDAR data: a 'primary building mask' and a 'secondary building mask'. Line segments are extracted from around the primary building mask, the segments around trees being removed using the normalized difference vegetation index derived from orthorectified multispectral images. 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. The proposed threshold-free evaluation system makes one-to-one correspondences using nearest centre distances between detected and reference buildings. A total of 15 indices are used to indicate object-based, pixel-based and geometric accuracy of the detected buildings. It is experimentally shown that the proposed technique can successfully detect rectilinear buildings, when assessed in terms of these indices including *completeness*, *correctness* and *quality*.

1 INTRODUCTION

Building detection from remotely sensed data has a wide range of applications including change detection, automatic city modeling, homeland security and disaster (flood or bush fire) management. Therefore, a large number of building detection techniques have been reported over the last few decades. These can be divided into three groups (Lee et al., 2008). The first group of algorithms uses 2D or 3D information from photogrammetric imagery (Mayer, 1999). These algorithms are complex due to involvement of detailed information in high-resolution images (Awrangjeb et al., 2010) and complicated and erroneous estimation of 3D (height) information (Sun et al., 2005). Algorithms in the second group consider building detection as a classification problem and detect building regions from LIDAR (LIght Detection And Ranging) data (Lee et al., 2008). However, the use of raw or interpolated data can influence the detection performance (Demir et al., 2009) resulting in poor horizontal accuracy for building edges (Yong and Huayi, 2008). As a result, it is hard to obtain a detailed and geometrically precise boundary using only LIDAR point clouds (Awrangjeb et al., 2010)

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). The third category of methods does use 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 (Lee et al., 2008). 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 (Yong and Huayi, 2008); only a few approaches with technical details have thus far been published (Rottensteiner et al., 2005).

In addition, 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), and the use of 1 to 2 evaluation indices only has characterized many studies (Demir et al., 2009, Vu et al., 2009). The majority of these (Rottensteiner et al., 2005, Rutzinger et al., 2009, Lee et al., 2008) 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 too subjective and likely to be controversial since there is no unique way to select the thresholds (Shufelt, 1999).

This paper aims at a successful integration of LIDAR data and photogrammetric imagery for building detection so that the improved detection performance is obtained. It also develops an automatic and threshold-free performance evaluation system using 15 indices from three categories: object-based, pixel-based and geometric. The performance of the proposed building detection approach has been assessed using the proposed evaluation system. Note that this paper is a condensed version of (Awrangjeb et al., 2010) with an extended experimental validation. It 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.

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 (Rottensteiner et al., 2005, Vu et al., 2009). As a result, they can suffer from poor horizontal accuracy for the detected buildings.

Secondly, there are integration techniques (Lee et al., 2008, Demir et al., 2009, Sohn and Dowman, 2007) which use both LIDAR data and imagery as the primary cues to delineate building outlines. They also employ imagery to remove vegetation. Consequently, they offer better horizontal accuracy for the detected

buildings. The proposed building detection technique falls into this group.

There are two groups of performance evaluation systems: 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). In (Rottensteiner et al., 2005) and (Rutzinger et al., 2009), a correspondence 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 (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*.

The evaluation systems can also be categorized into pixel-based (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.

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 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.

3.1 Generation of Masks

The raw LIDAR data is divided into groups where each group corresponds to a tile of 450×450 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 those of large heights 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 (Digital Elevation Model) 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 $T_h = H_g + 3.5m$ is applied to split the LIDAR points into two sets: points with lower heights and points with higher

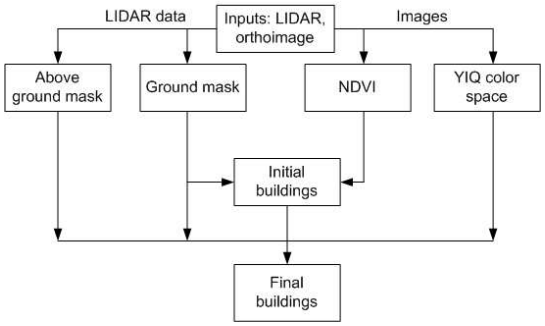

Figure 1: Flow diagram of the proposed building detection technique.

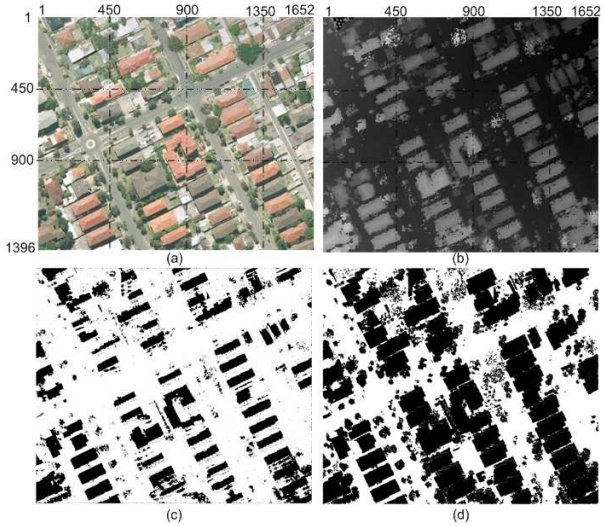

Figure 2: (a) A test scene, (b) LIDAR data (shown in gray-scale), (c) primary building mask and (d) secondary building mask.

heights. The first set marks white for each of its point in the primary building mask M_p which is initially a completely black mask. The second set marks black for each of its point 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.

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 (Awrangjeb et al., 2009). 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 sigma of the NDVI value Υ is calculated on both sides of each line segment. If it is above a threshold $T_{ndvi} = 64$ 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 the 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 θ_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 θ_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 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 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 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.

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.

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

4 PROPOSED EVALUATION SYSTEM

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. 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.

4.1 Detected and Reference Building Sets

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 detection technique. Each entity $b_{d,i}$ is an array of four vertices and the centre 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 a 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 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.

4.2 Overlapping Sets

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. 4(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. 4(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. 4(b)) or constitute combinations of true and false detections (Figs. 4(c)-(e)).

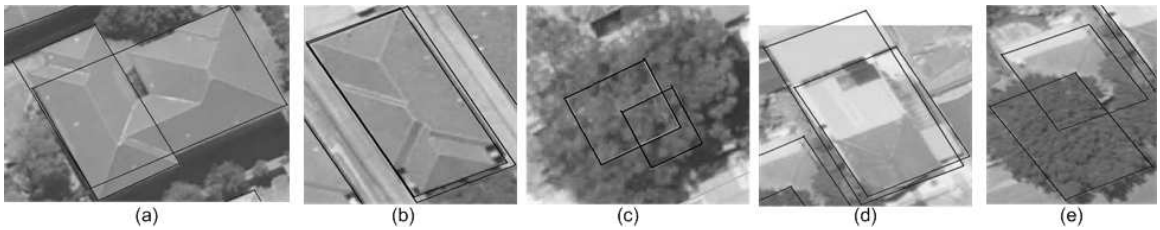

Figure 4: Different types of detection overlaps: (a) natural, (b) multiple detection, (c) false-false, (d) true-true and (e) true-false.

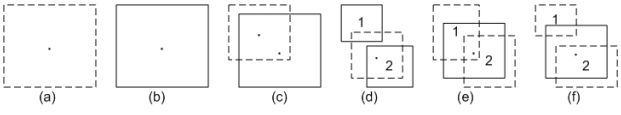

Figure 5: Different situations while establishing pseudo one-to-one correspondence. Solid rectangles denote reference entity and dotted rectangles denote detection entity. The center of each entity is either denoted by a dot or by a number.

Two overlapping sets are obtained from B_d and B_r to facilitate establishing one-to-one correspondences. One is *detection overlapping set* $O_d = \{o_{d,i}\}$, where $0 \leq i \leq m$, of B_d with respect to B_r and another is *reference overlapping set* $O_r = \{o_{r,j}\}$, where $0 \leq j \leq n$, of B_r with respect to B_d. Each entity $o_{d,i}$ in O_d contains a list of entities from B_r which $b_{d,i}$ overlaps. If $b_{d,i}$ overlaps none from B_r, $o_{d,i}$ is empty. Similarly, each entity $o_{r,j}$ in O_r contains a list of entities from B_d which $b_{r,j}$ overlaps.

In order to obtain O_d, a total of 16 points are considered on each rectangular entity $b_{r,j}$ in B_r: 4 vertices and 3 points on each side at equal distant. All the entities $b_{r,j}$ in B_r are tested against each entity $b_{d,i}$ in B_d. If at least 1 out of 16 points of $b_{r,j}$ falls inside $b_{d,i}$, $b_{d,i}$ overlaps $b_{r,j}$. All $b_{r,j}$ which overlap each $b_{d,i}$ are included into $o_{d,i}$. O_r is obtained following the same procedure as that above.

4.3 Pseudo One-to-One Correspondences

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. *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.

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. As shown in Fig. 4(b) there may be more than one detection of the same building. In order to establish the one-to-one correspondences it is important that only one of these detections is considered as a TP. The rests are counted as MDs and a new index named *multiple detection rate* is defined. Note that MD is counted for the detection set only and there will be no one-to-one correspondence for an MD.

The iterative procedure below establishes the pseudo one-to-one correspondences between the detection and reference sets.

1. If the overlapping entity $o_{d,i}$ corresponding to a detection entity $b_{d,i}$ is empty, then $b_{d,i}$ is marked as an FP (Fig. 5(a)). Similarly, if the overlapping entity $o_{r,j}$ corresponding to a reference entity $b_{r,j}$ is empty, then $b_{r,j}$ is marked as an FN (Fig. 5(b)).

2. For each overlapping entity $o_{d,i}$ corresponding to a detection entity $b_{d,i}$, if $b_{d,i}$ has not been marked yet: suppose the entities in $o_{d,i}$ are sorted as $\{b_{r,j_1}, b_{r,j_2}, ..., b_{r,j_k}\}$ $(k \geq 1)$ in the ascending order of their center distances to the center of $b_{d,i}$. This means b_{r,j_1} is the closest overlapped reference entity to $b_{d,i}$. Further suppose the entities in o_{r,j_1} are sorted as $\{b_{d,i_1}, b_{d,i_2}, ..., b_{d,i_l}\}$ $(l \geq 1)$ in the ascending order of their center distances to the center of b_{r,j_1}. If b_{d,i_1} and $b_{d,i}$ are the same entity, this means $b_{d,i}$ is the closest overlapped detection entity to b_{r,j_1}. In this case, the following steps are followed:

   - Establish a one-to-one correspondence between $b_{d,i}$ and b_{r,j_1} by marking both of them as TPs (Fig. 5(c)).
   - For each of the remaining entities b_{r,j_s} $(2 \leq s \leq k)$ in $o_{d,i}$, $b_{d,i}$ is removed from the overlapping entity o_{r,j_s}. In Fig. 5(d), since based on center distances reference 2 is more close to the detection entity than reference 1, reference 2 is a TP and reference 1 is an FP.
   - For each of the remaining entities b_{d,i_t} $(2 \leq t \leq l)$ in o_{r,j_1}, if the overlapping entity o_{d,i_t} of b_{d,i_t} contains only one reference entity (which is obviously b_{r,j_1}) it is checked whether b_{d,i_t} and $b_{d,i}$ overlap each other. If they overlap each other, then b_{d,i_t} is marked as an MD (Fig. 5(e)). If they do not overlap each other, then b_{r,j_1} is removed from o_{d,i_t} which becomes empty immediately (Fig. 5(f)). Otherwise, if o_{d,i_t} contains more than one reference entities (including b_{r,j_1}), then b_{r,j_1} is removed from o_{d,i_t}.

The above procedure continues until all the detection and reference entities are marked. Note that any of the overlapping entity which becomes empty in Step 2 of any iteration, the corresponding detection or reference entity will be marked (as an FP or FN) in Step 1 of the next iteration. Since in practice, in most cases there will be only one overlap for each entity, the above iterative procedure converges quickly after a few iterations.

4.4 Evaluation Indices

Seven indices are used for object-based evaluation. *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). *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 entities. *Reference cross-lap rate* is defined as the percentage of reference

entities which are overlapped by more than one detected entity (see (Awrangjeb et al., 2010) for formal definitions).

A total of 7 pixel-based evaluation indices are also 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).

Root-mean-square-error (RMSE) values (Song and Haithcoat, 2005) 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.

5 PERFORMANCE STUDY

5.1 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. 2(a)), the second covering an area of $155m \times 219m$ (Fig. 6(a)), the third covering an area of $228m \times 189m$ (Fig. 6(b)) and the fourth covering an area of $586m \times 415m$ (Fig. 6(c)). Last-pulse LIDAR data with a point spacing of $0.5m$ was used. 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 R-G-B are in the order of IR-Red-Green in order to be used in the standard NDVI formula.

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 (BaristaSoftware, 2010). All rectangular structures, recognizable as buildings and above the height threshold T_h were digitized. The reference data included garden sheds, garages, etc., that were sometimes as small as $10m^2$ in area. Altogether, 70, 62, 60 and 370 buildings from the four test scenes formed the reference sets.

5.2 Results and Discussion

Table 1 shows the object-based evaluation results and Table 2 shows the pixel-based evaluation results. The geometric accuracy (RMSE) for three scenes was $1.99m$, $1.95m$, $1.95m$ and $2.38m$ with an average accuracy of 14.5 pixels ($2.17m$).

In object-based evaluation, more than 94% *completeness* and *correctness* resulted in an average 91% *quality* with at least 5% 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 81% of building areas were completely detected, resulting in a 19% *omission error*, 88% of detected areas were correct, offering 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 proposed technique is lower than its *false negative rate*.

Figure 6: Detected buildings on the orthoimages.

Overall, both in object- and pixel-based evaluations, the proposed detection technique performed better on Scene 1 than on Scene 2 in terms of all indices except *multiple detection rate* and *detection overlap rate*. There were two reasons for this: a) some buildings were detected twice in Scene 1, and b) though in Scene 1 all true buildings were detected, in Scene 2 some false buildings (actually trees) were detected and some true building parts were missed. 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. Almost the same was observed for Scene 4. In the geometric evaluation, in terms of RMSE, there was about $0.4m$ worse positional accuracy for Scene 4 than the other three scenes.

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* 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, (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).

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	98.6	97.2	95.9	4.1	5.4	1.4	5.7
Scene 2	95.2	95.2	90.8	3.1	3.1	1.6	6.5
Scene 3	98.3	92.2	90.8	4.5	9.0	13.4	23.3
Scene 4	95.1	95.1	90.7	6.1	18.3	17.5	28.7
Average	**95.9**	**94.7**	**91.4**	**5.1**	**12.5**	**12.5**	**21.7**

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	78.5	89.0	71.5	21.6	10.7	12.3	27.5
Scene 2	77.7	87.4	69.8	22.3	12.3	14.5	28.8
Scene 3	80.5	91.4	74.8	19.5	8.3	9.5	24.3
Scene 4	81.4	85.1	71.3	18.6	14.1	17.5	22.9
Average	**80.4**	**87.5**	**72.0**	**19.7**	**12.0**	**14.5**	**24.6**

6 CONCLUSIONS AND FUTURE WORK

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 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* 95.9%) is high when compared to pixel-based *completeness* (*matching overlay* 81.4%). However, the geometric positional accuracy remains relatively poor (14.5 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 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.

REFERENCES

Awrangjeb, M., Lu, G., Fraser, C. S. and 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.

Awrangjeb, M., Ravanbakhsh, M. and Fraser, C. S., 2010. Automatic detection of residential buildings using lidar data and multispectral imagery. ISPRS Journal of Photogrammetry and Remote Sensing.

BaristaSoftware, 2010. www.baristasoftware.com.au.

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

Demir, N., Poli, D. and Baltsavias, E., 2009. 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), pp. 71–76.

Foody, G., 2002. Status of land cover classification accuracy assessment. Remote Sensing of Environment 80(1), pp. 185–201.

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

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

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

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

Rottensteiner, F., Trinder, J., Clode, S. and 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), pp. 135–149.

Rutzinger, M., Rottensteiner, F. and Pfeifer, N., 2009. 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), pp. 11–20.

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

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

Sohn, G. and Dowman, I., 2007. Data fusion of high-resolution satellite imagery and lidar data for automatic building extraction. ISPRS Journal of Photogrammetry and Remote Sensing 62(1), pp. 43–63.

Song, W. and Haithcoat, T., 2005. Development of comprehensive accuracy assessment indexes for building footprint extraction. IEEE Transactions on Geoscience and Remote Sensing 43(2), pp. 402–404.

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

Vu, T., Yamazaki, F. and Matsuoka, M., 2009. Multi-scale solution for building extraction from lidar and image data. International Journal of Applied Earth Observation and Geoinformation 11(4), pp. 281–289.

Yong, L. and Huayi, W., 2008. 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), pp. 197–202.