

A NEW MASK FOR AUTOMATIC BUILDING DETECTION FROM HIGH DENSITY POINT CLOUD DATA AND MULTISPECTRAL IMAGERY

Mohammad Awrangjeb*, Fasahat U. Siddiqui

Institute for Integrated and Intelligent Systems, Griffith University, Nathan, QLD 4111, Australia

Phone: +61 7 373 55032, E-mail: m.awrangjeb@griffith.edu.au, f.siddiqui@griffith.edu.au

Web: www.ict.griffith.edu.au/awrangjeb

KEY WORDS: Building extraction, building mask, building detection, point cloud data, LIDAR, multispectral imagery

ABSTRACT:

In complex urban and residential areas, there are buildings which are not only connected with and/or close to one another but also partially occluded by their surrounding vegetation. Moreover, there may be buildings whose roofs are made of transparent materials. In transparent buildings, there are point returns from both the ground (or materials inside the buildings) and the rooftop. These issues confuse the previously proposed building masks which are generated from either ground points or non-ground points. The normalised digital surface model (nDSM) is generated from the non-ground points and usually it is hard to find individual buildings and trees using the nDSM. In contrast, the primary building mask is produced using the ground points, thereby it misses the transparent rooftops. This paper proposes a new building mask based on the non-ground points. The dominant directions of non-ground lines extracted from the multispectral imagery are estimated. A dummy grid with the target mask resolution is rotated at each dominant direction to obtain the corresponding height values from the non-ground points. Three sub-masks are then generated from the height grid by estimating the gradient function. Two of these sub-masks capture planar surfaces whose height remain constant in along and across the dominant direction, respectively. The third sub-mask contains only the flat surfaces where the height (ideally) remains constant in all directions. All the sub-masks generated in all estimated dominant directions are combined to produce the candidate building mask. Although the application of the gradient function helps in removal of most of the vegetation, the final building mask is obtained through removal of planar vegetation, if any, and tiny isolated false candidates. Experimental results on three Australian data sets show that the proposed method can successfully remove vegetation, thereby separate buildings from occluding vegetation and detect buildings with transparent roof materials. While compared to existing building detection techniques, the proposed technique offers higher object-based completeness, correctness and quality, specially in complex scenes with aforementioned issues. It is not only capable of detecting transparent buildings, but also small garden sheds which are sometimes as small as 5 m^2 in area.

1. INTRODUCTION

Automatic building detection from remote sensing data has various applications including urban planning and disaster (e.g., bushfire) management. In the literature, a number of building detection techniques have been reported over the last couple of decades. These can be divided into three major groups (Lee et al., 2008). Firstly, there are many algorithms which use 2D or 3D information from photogrammetric imagery. Secondly, there have been several attempts to detect building regions from LIDAR (LIght Detection And Ranging) data. Finally, since LIDAR and imagery each have particular advantages and disadvantages in horizontal and vertical positioning resolution and accuracy, several authors have promoted an integration of LIDAR data and imagery as a means of advancing building detection. More specifically, intensity and height information from LIDAR can be used with texture and region boundary information from imagery to improve detection accuracy (Habib et al., 2010). However, the success of automatic building detection is still largely impeded by scene complexity, incomplete cue extraction and sensor dependency of data (Sohn and Dowman, 2007). Vegetation, and especially trees, can be the prime cause of scene complexity and incomplete cue extraction. Important building cues can be completely or partially missed due to occlusions and shadowing from trees (Awrangjeb et al., 2012).

Majority of the existing building detection techniques, which in-

volve LIDAR data, are based on the two previously proposed building masks: the normalised digital surface model (nDSM) (Rottensteiner et al., 2005) and primary mask (Awrangjeb et al., 2010). The nDSM is generated from the non-ground points and usually it connects neighbouring buildings and trees. Consequently, it is hard to find individual buildings and trees using the nDSM. In contrast, the primary building mask is produced using the ground points. Since the ground points represent the ground in this mask, due to ground points returned from the transparent building floors, such buildings are not captured in this mask.

This paper proposes a new building mask for automatic building detection from LIDAR data and multispectral imagery. First, objects below a given threshold above the ground, such as bushes, cars and carports, are removed from the raw LIDAR data. Second, the dominant directions of non-ground lines extracted from the multispectral imagery are estimated. Third, a dummy grid with the target mask resolution is rotated at each dominant direction to obtain the height values from the non-ground points. Fourth, three sub-masks are then generated from the height grid by estimating the gradient function. Fifth, all the sub-masks generated in all estimated dominant directions are combined to produce the candidate building mask. Finally, the final building mask is obtained through removal of planar vegetation, if any, and tiny isolated false candidates. In experimentation, a higher object-based performance has been observed when tested on three Australian data sets, specially in complex scenes with dense vegetation causing shadow and occlusions.

*Corresponding author

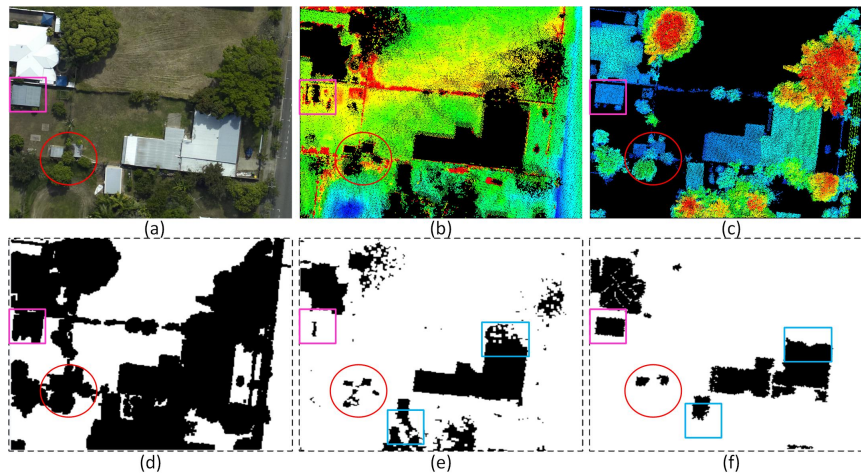


Figure 1. A complex scene: (a) aerial image, (b) ground points, (c) non-ground points, (d) normalised-DSM generated from the non-ground points, (e) primary mask generated from ground points and (f) proposed building mask generated from non-ground points.

2. SEARCH FOR A NEW BUILDING MASK

Figure 1 shows a complex scene, where buildings are not only connected with and partially occluded by their surrounding vegetation but also they are close to each other. Moreover, there is a building whose roof is made of transparent materials (shown within the pink coloured rectangle). Consequently, for this building there exist both ground and non-ground points (see points within the pink coloured rectangle in Figures 1(b)-(c)). These two phenomena confuse any building masks which are generated from the ground or the non-ground point sets.

Before generating any masks, the nDSM (Rottensteiner et al., 2005) and primary mask (Awrangjeb et al., 2010) apply a height threshold T_h (e.g., the ground height plus 1 m) to the input point-cloud data in order to divide the input data into ground and non-ground point sets. The nDSM, shown in Figure 1(d), is generated using the non-ground point set in Figure 1(c) and represents non-zero heights indicating the elevated objects above T_h . As can be seen, buildings and trees are found connected with one another in the nDSM. Thus, it is hard to find the individual buildings using the nDSM. In contrast, the primary mask, shown in Figure 1(e), is generated using the ground point set in Figure 1(b). In this mask, buildings and trees are found thinner than those in the nDSM. Consequently, individual buildings and vegetation are easily recognisable in the primary mask. However, it has two main problems. Firstly, some vegetation are still found connected with the neighbouring trees (see light blue coloured rectangles in Figure 1(e)). Secondly, the transparent buildings within the pink coloured rectangle is almost missed.

In order to eliminate the aforementioned problems, a new building mask based on the non-ground points is proposed in this paper. As shown in Figure 1(f), the new building mask has the following advantages. Firstly, any vegetation in the scene is mostly removed. Thus, buildings become easily identifiable. Secondly, buildings with transparent roof materials are detected. Finally, it also detects small garden sheds which are sometimes as small as 5 m^2 in area.

3. PROPOSED BUILDING DETECTION TECHNIQUE

Figure 2 shows the flow diagram for the proposed new building detection technique. In the following subsections, each step is described in detail.

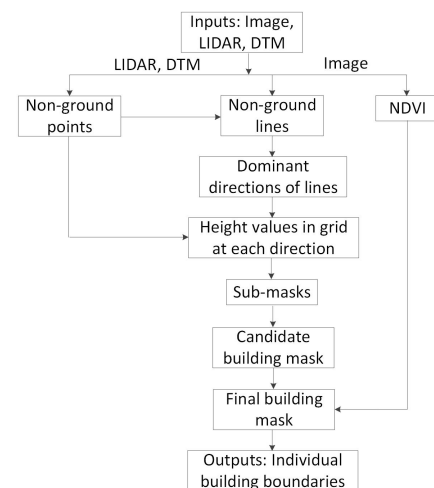


Figure 2. Flow diagram for new building detection technique.

3.1 Find Non-ground Points

If a Digital Terrain Model (DTM) is not available, one can be generated from the input LIDAR point cloud data using a commercial software. We assume that the DTM is given as an input to the proposed technique. Then, for each LIDAR point, the corresponding DTM height is used as the ground height, H_g . A height threshold $T_h = H_g + h_c$, where h_c is a height constant that separates low height objects from high height objects, is then applied to the LIDAR data (Awrangjeb et al., 2010). Thus, the point cloud is divided into two groups: ground and non-ground points. The second group consists of the points that represent elevated objects, such as buildings and trees with heights above T_h . The value of h_c was set at the minimum building heights (2.5 m) in earlier studies (Rottensteiner et al., 2005, Awrangjeb et al., 2010). However, in order to extract garden sheds, which are sometimes may have lower heights than buildings, $h_c = 1 \text{ m}$ (Awrangjeb and Fraser, 2014) has been set in this study. Figure 3(b) shows the non-ground points obtained from the input sample LIDAR data.

3.2 Find Non-ground Image Lines

To extract straight lines from the input multi-spectral image the procedure introduced in (Awrangjeb et al., 2010) is followed. The

image is first converted into a grayscale image and the Canny edge algorithm is applied to find edges from the grayscale image. Corners are then detected on the extracted edges. The detected corners on each extracted edge show the significant changes in direction of the edge. Two end points of an edge are also considered corners. A straight line is fitted between the two consecutive corners on each extracted edge. Finally, extracted straight lines are divided into ground and non-ground lines using the non-ground LIDAR points in Figure 3(b). If there are non-ground points on either side of a line it is decided to be a non-ground image line. Hence, a non-ground line can be on a building or on a tree. Figure 3(c) shows the ground lines in red colour and the non-ground lines in cyan colour. The maximum point-to-point distance d_m in the input point cloud data is used to look for neighbouring non-ground points for a line and lines with length less than the minimum building width $W_m = 3$ m are removed (Rottensteiner et al., 2005). The latter parameter removes many small lines obtained on the vegetation.

3.3 Obtain Dominant Directions of Lines

Assuming buildings in a given scene are rectilinear in nature, a histogram analysis is proposed to automatically obtain all dominant directions of lines, which are extracted mainly from buildings. Such dominant line directions come into pairs and in each pair (ϕ, φ) two directions are orthogonal to each other. A simple rectangular building has two dominant directions ϕ and φ : ϕ is along the length of the building and φ is along the width. The ridge lines on a rectilinear building usually seem to be diagonal (at 45°) to one of the dominant directions. In an input scene, buildings may be of complex rectilinear shapes and they may be oriented in different directions. There may also be lines extracted from trees. That means there may be more than one pair of dominant directions of lines in the input scene.

In the proposed histogram analysis, the tangent angle (slope) θ of each non-ground line in the scene is converted in the range $[0^\circ, 180^\circ]$. A direction histogram of 32 bins (bin distance $\theta_d = 5.625^\circ = \frac{\pi}{32}$ rad) is then used to divide the non-ground image lines based on their values of θ . The histogram is circular in the sense that the first and last bins are considered neighbours of each other (i.e., consecutive bins). Figure 3(d) shows all 8 bins, where at least one long line (minimum $W_m = 3$ m in length) is found in a bin, in different colours. The mean of slope for lines in each bin is measured. The bins are then sorted based on the number of lines.

A selection procedure is now followed based on the trivial observation that on rectilinear buildings edge and ridge lines are mostly parallel, perpendicular or diagonal to one of the dominant directions. Starting for the largest bin b_L , if b_L is parallel, perpendicular or diagonal to any other bins b_a , then b_L is marked to be a new dominant direction. If b_a is parallel to b_L , then b_a is marked not to consider anymore as b_L represents its direction. This parallel case happens when a dominant direction fall in and around the border of two neighbouring bins. If b_a is perpendicular to b_L , then b_a is marked to form a pair of dominant directions with b_L . If b_a is diagonal to b_L , then b_a is also marked as a new dominant direction and the bin that is orthogonal to b_a is searched and marked to form another pair of dominant directions. At this point if no such b_a exists unmarked, the next largest bin which still remains unmarked is considered to be b_L and the loop continues. If in an iteration no bin is found to be parallel, perpendicular or diagonal to b_L , then b_L is marked as a noise (i.e., not a dominant direction of lines).

Note that some of the dominant directions of lines as obtained above may not be the actual dominant directions of buildings, but they are dominant directions of those lines which are mainly ridge lines of buildings.

For above parallel, perpendicular and diagonal tests θ_d is used as a threshold, which will allow to have dominant directions in a given scene at an interval of $11.25^\circ = \frac{\pi}{16}$ rad (i.e., maximum 8 pairs of dominant directions). Depending on the test scene $\theta_d = 11.25^\circ$ can be used for parallelism test in order to avoid additional dominant directions which can happen if a true orthophoto is not used.

Figure 3(e) shows the lines in two bins for which two actual dominant directions are estimated: the mean slope of the cyan coloured lines is 173.6° and that of the yellow is 124.3° . All the parallel and perpendicular lines to these two sets (shown in cyan and yellow colours) of lines are not shown in Figure 3(e), but all lines in bins are shown in Figure 3(d). The red line in Figure 3(e) shows a ground line, but has been classified as a non-ground line due to LIDAR points on the cloth hoist. However, it was ultimately removed as a non-dominant direction because it does not have any parallel, perpendicular or diagonal lines. Let the set of first dominant directions be $\{\phi_i\}$, where $i \geq 1$, and $\{\varphi_i\}$ can be easily calculated from $\{\phi_i\}$, if necessary. For the sample scene, $\phi_1 = 173.6^\circ$ and $\phi_2 = 124.3^\circ$.

3.4 Generate Mask

The generation of building mask M_b takes place in two main steps. Firstly, a sub-mask $M_{s,i}$ is generated at each dominant direction of lines ϕ_i and secondly, sub-masks from all ϕ_i are combined to produce a single building mask M_b .

Let the size of M_b be $h_b \times w_b$, which is the same as the size of the input scene and defined by the bottom-left and top-right corners of the bounding rectangle that contains all input LIDAR points. The resolution of M_b is kept fixed at $r_b = 0.25$ m (Awrangjeb and Fraser, 2014). Each $M_{s,i}$ has the same size and resolution as M_b .

3.4.1 Generate Sub-masks Each sub-mask $M_{s,i}$ is generated by rotating it to a dominant direction of lines ϕ_i with respect to the centre of the mask. However, this rotation will cause some parts of the input scene being uncovered by the rotated mask. In order to cover the whole scene, a temporary sub-mask M_{ts} is used. It is a square mask with each side equalling to $h_{bt} = w_{bt} = \sqrt{2} \max(h_b, w_b)$. M_{ts} covers the whole scene (i.e., M_b or $M_{s,i}$), even when M_{ts} is rotated at any rotation angle. Figure 4(a) shows that when M_{ts} is rotated at 45° it covers the whole scene area.

Consequently, M_{ts} is rotated, instead of $M_{s,i}$, at each dominant angle $\{\phi_i\}$ to produce a temporary sub-mask $M_{ts,i}$. For example, for $i = 1$ Figure 3(f) shows $M_{ts,1}$ when rotated at the first dominant direction ($\phi_1 = 173.6^\circ$). The mask is shown for visualisation only and cropped to fit the figure. After rotation, the corresponding height value in each pixel location of $M_{ts,1}$ is assigned from the non-ground point set. Gradients are then calculated in along and across the dominant direction. Considering the error in the supplied point cloud data a threshold $T_p = 0.15$ m is applied to convert each gradient image into a binary mask. Let these two gradient (binary) images be $F_{x,1}$ and $F_{y,1}$ shown in Figure 3(g)-(h). Basically, $F_{x,1}$ represents planes where the height (ideally) remains constant along ϕ_1 but changes along φ_1 .

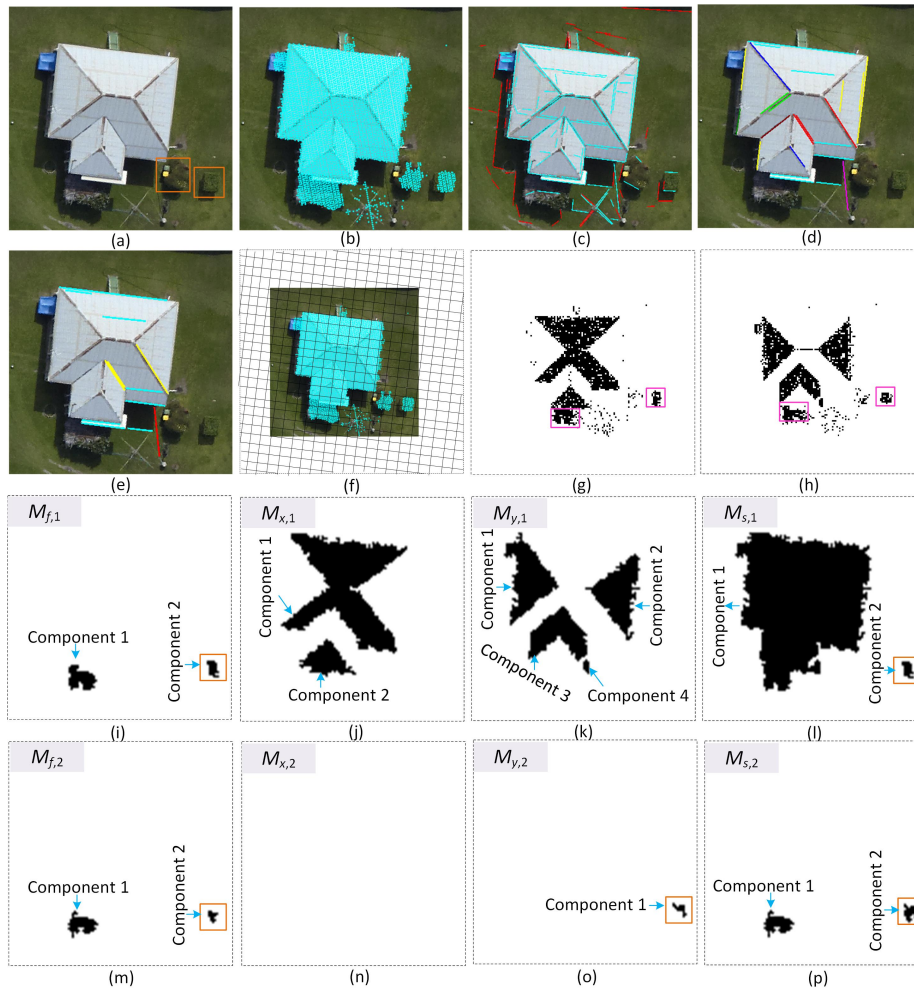


Figure 3. Proposed mask generation from a sample scene: (a) aerial image, (b) non-ground points, (c) all image lines (red for ground and cyan for non-ground lines), (d) long lines in eight histogram bins are shown in different colours (lines are at least 3 m in length), (e) lines in two bins (cyan and yellow) for which two actual dominant directions are estimated (red line shows a non-dominant direction), (f) temporary sub-mask $M_{ts,i}$ rotated at the first dominant direction ($\phi_1 = 173.6^\circ$), (g) gradient image $F_{x,1}$, (h) gradient image $F_{y,1}$, (i) flat areas $M_{f,1}$, (j) sloppy areas $M_{x,1}$ along ϕ_1 , (k) sloppy areas $M_{y,1}$ across ϕ_1 , (l) sub-masks $M_{s,1}$, (m) flat areas $M_{f,2}$, (n) sloppy areas $M_{x,2}$ along ϕ_2 , (o) sloppy areas $M_{y,2}$ across ϕ_2 , (p) sub-masks $M_{s,2}$.

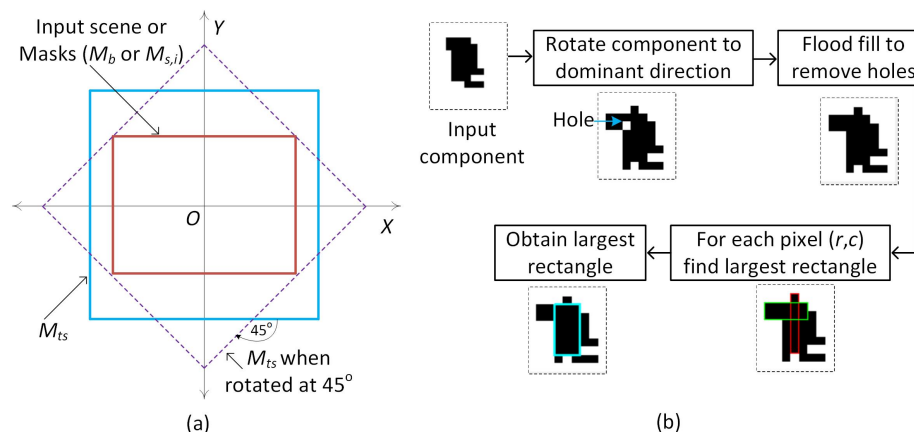


Figure 4. (a) Generation of the sub-mask – defining building mask M_b , sub-mask $M_{s,i}$ and temporary sub-mask M_{ts} . M_b and $M_{s,i}$ have the same size as the input scene, (b) Steps to find the largest rectangle within a component.

$F_{y,1}$ represents planes where the height remains constant along φ_1 but changes along ϕ_1 . Both of them also contain the flat surfaces, as shown within pink coloured rectangles in Figure 3(g)-

(h), where the height (ideally) remain constant in all directions. $F_{x,1}$ and $F_{y,1}$ are then converted back to the original orientation and size. In addition, occasional small holes, which may exist due

to height jumps between planes or noise in the data, in $F_{x,1}$ and $F_{y,1}$ are filled and small isolated components (maximum 1 m^2 , i.e., 16 pixels) are removed. Simple binary operations (AND, OR, SUBTRACTION, etc.) are applied to separate the flat areas from $F_{x,1}$ and $F_{y,1}$. Let these three binary images be $M_{x,1}$ (after flat areas are removed from $F_{x,1}$), $M_{y,1}$ (after flat areas are removed from $F_{y,1}$) and $M_{f,1}$ (contains all flat areas). Figures 3(i)-(k) show all these three images. One of the areas in the flat image is a cuboid shaped vegetation with a flat top (shown within orange coloured rectangle in Figure 3(i)). Figure 3(l) shows $M_{s,1}$ which is simply the result of the OR operation among $M_{x,1}$, $M_{y,1}$ and $M_{f,1}$.

Figures 3(m)-(p) show all the three images and the sub-mask for $\phi_2 = 124.3^\circ$. Since, ϕ_2 is not a principal direction of buildings, but one of the dominant angles for non-ground lines, the building area is mostly missed in the sub-mask. However, as expected the flat areas, as shown Figure 3(m), are clearly captured. While $M_{x,2}$ does not contain any considerable area, $M_{y,2}$ contains a part of the vegetation (shown within orange coloured rectangle in Figure 3(o)).

Figure 3(a) shows two separate vegetation within orange coloured rectangles. As shown in two sub-masks in Figures 3(l) and (p), one of the vegetation is totally removed as it has height variation on its top. The other could not be removed at this point as it is cut into a cuboid with flat top.

3.4.2 Combine Sub-masks Since, it is impossible to distinguish individual components in each $M_{ts,i}$ (at dominant direction ϕ_i), the three binary images ($M_{x,i}$, $M_{y,i}$ and $M_{f,i}$) are used to combine the sub-masks.

For each binary image, a connected component analysis is conducted. For each connected component the MATLAB function *bwconncomp* provides its area, convex hull and pixel locations. Considering that the minimum dimension (length or width) of a single isolated building is $W_m = 3 \text{ m}$, any component which is smaller than $A_m = 9 \text{ m}^2$ is subject to a dimensionality test in order to keep the component as a part of a big building. The largest rectangle R_l containing only the pixels of the component is obtained. Both dimensions of R_l should be at least $l_m = 1 \text{ m}$ in the test. This means the minimum width or length of a plane is 1 m . If a component fails this test it is removed as a vegetation.

Figure 4(b) shows the steps for finding R_l for the component shown within orange coloured rectangle in Figure 3(i). The output for each step is shown at the bottom of that step. First, the component is rotated to its corresponding dominant direction ϕ_i ($i = 1$ for this component). For rotation, the procedure described in Section 3.4.1 is followed. Second, the rotated component is flood filled (using the MATLAB function *imfill*) in order to remove any holes that exist due to rotation. Thirdly, for each pixel (r, c) find the height and width of the largest all-black rectangle with its upper-left corner at (r, c) . Two such rectangles are shown in red and green colours. Finally, from all such rectangles the largest rectangle R_l , shown in cyan colour, is obtained.

Thereafter, the overlap between any two components at any dominant directions are examined. There may be a small thin overlap along the edges between neighboring components due to the digitisation effect. Moreover, when many dominant directions exist in a given scene, some building planes can be captured in more than one dominant direction. In some cases, even neighbouring (near) parallel planes can be found partially or fully merged in

some dominant directions. For making sure that each connected component represents one or more building planes and no two components represent the same plane, the following procedure is followed in order to select, remove and/or replace the connected components.

All the connected components ζ_j , where $j \geq 1$, are initially marked as ‘selected’ irrespective of their dominant directions. They are then sorted in descending order of their areas. Starting from the largest ‘selected’ component ζ_m , all components $\zeta_{n,k}$, where $k \geq 1$, within the neighbourhood of ζ_m are found. This neighbourhood search is simply accomplished considering the intersection of convex hulls. Now, for each neighbour $\zeta_{n,k}$ its percentage of overlap with ζ_m is estimated. If there is a major overlap (more than 50%), it is tested that the overlap area has at least one component which is at least 1 m^2 in area (*Test 1*). This test ensures that thin overlaps along the edges of neighbouring planes does not cause any wrong decision. If *Test 1* is passed, large components from the non-overlap area of this $\zeta_{n,k}$ are obtained separately and marked ‘selected’ (*Test 2*). To find a large area, again the largest rectangle R_l containing only the pixels of the component is obtained. Both dimensions of R_l should be at least $l_m = 1 \text{ m}$ in the test. If *Test 1* is failed, a slightly flexible condition is applied to still consider the non-overlap area of $\zeta_{n,k}$ if it has large components (*Test 3*). These large components should be at least 1 m^2 in area. If at least one large area is found (*Test 2* or *3* is passed), the corresponding $\zeta_{n,k}$ is marked as ‘replaced’. If no such large area is found (*Test 2* or *3* is failed), then $\zeta_{n,k}$ is marked as ‘removed’. Note that in the above procedure, when necessary, the connected component analysis is conducted on the overlap and non-overlap areas separately for each $\zeta_{n,k}$.

For the sixth largest component $\zeta_{m,6}$, which is Component 1 in $M_{f,1}$ (see Figure 3(i)), two components are found in its neighbourhood: $\zeta_{n,1}$, which is Component 2 in $M_{x,1}$ (see Figure 3(j)), and $(\zeta_{n,2})$, which is Component 1 in $M_{f,2}$ (see Figure 3(m)). No overlap exists between $\zeta_{n,1}$ and $\zeta_{m,6}$. Figure 5(a) shows the overlap area between $\zeta_{m,6}$ and $\zeta_{n,2}$ and the non-overlap area for $\zeta_{n,2}$. Since the overlap area is 93% and 5.4 m^2 in area (pink area), *Test 1* is passed. *Test 2* is conducted but is failed, because six components (blue areas), each less than 1 m^2 in area, are found. Consequently, $\zeta_{n,2}$ is marked as ‘removed’.

Figure 5(b) shows individual surviving components from $M_{x,i}$, $M_{y,i}$ and $M_{f,i}$ in green, blue and red colours respectively. All these components which still remain ‘selected’ are combined (binary OR operation) to find the building mask. The combined mask is shown in Figure 5(c). Thereafter, the combined mask is flood filled in order to fill any holes that may exist due to digitisation effect and slope discontinuity between planes. A morphological opening filter is applied with a square structural element of 1 m^2 in size to reduce the zigzag pattern around the building boundary. Figure 5(d) shows the binary mask, where there are two candidate buildings exist.

3.5 Finalise Buildings

As shown in Figures 5(c-d) within the orange coloured rectangle, a small but dense vegetation could not be removed due to its planar top. For such a dense and planar vegetation the laser ray cannot penetrate through to the trunk or the ground and, thus, the vegetation show low gradient values and eventually it remains as a building in the building mask. However, such vegetation (false buildings) are rare and small in size.

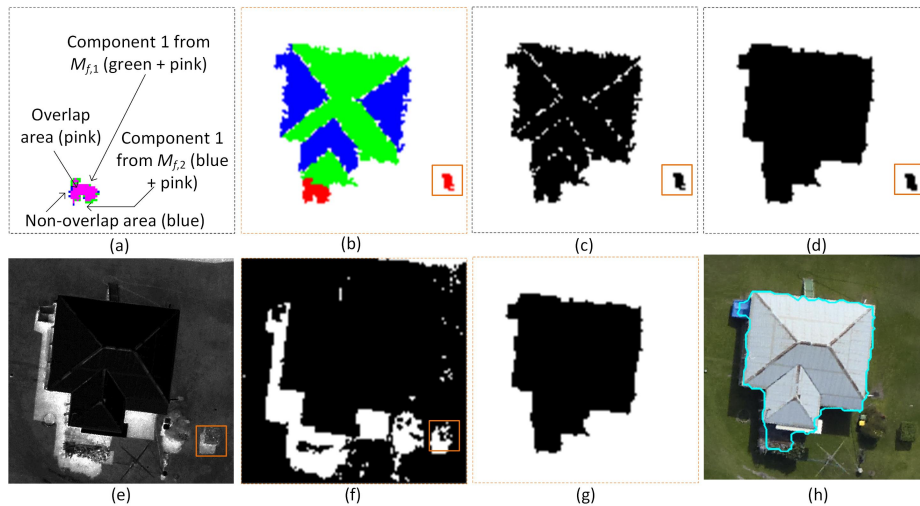


Figure 5. Combining components of sub-masks: (a) analysing overlap and non-overlap areas between components, (b) coloured mask, (c) binary mask, (d) candidate buildings, (e) normalised difference vegetation index (NDVI) for sample scene in Figure 1(a), (f) NDVI mask, (g) final building mask and (h) final building boundaries.

In order to remove such vegetation, a simple analysis with normalised difference vegetation index (NDVI) is followed (Awrangjeb et al., 2010) for candidate buildings smaller than 10 m^2 . Figure 5(e) shows the NDVI image for the sample scene in Figure 1(a). Figure 5(b) shows the binary NDVI mask after application of the NDVI threshold (see (Awrangjeb et al., 2010) for details). If within a candidate building the mean NDVI exceeds the NDVI threshold then this candidate is removed as a vegetation. At last, small (say, 1 m^2 in area) isolated buildings are removed. Figure 5(g) shows the final building mask and Figure 5(h) shows the building boundary. The boundary extraction procedure in (Awrangjeb and Fraser, 2014) is followed.

4. PERFORMANCE STUDY

In the performance study conducted to assess the proposed approach, three data sets from two different areas were employed. The objective evaluation followed a previously proposed automatic and threshold-free evaluation system (Awrangjeb and Fraser, 2012).

The planimetric accuracy has been measured in both object and image space. In object-based evaluation the number of buildings has been considered – whether a given building in the reference set is present in the detection set. Five indices are used for object-based evaluation: completeness C_m , correctness C_r , quality Q_l , detection cross-lap C_{rd} (over-segmentation error) and reference cross-lap C_{rr} (under-segmentation error) rates. In pixel-based image space evaluation the number of pixels has been considered – whether a pixel in a reference building is present in any of the detected buildings. A total of 5 pixel-based evaluation indices are used, these being: completeness C_{mp} , correctness C_{rp} , quality Q_{lp} , branching factor B_f and miss factor M_f . In addition, root-mean-square-error (RMSE) has been used to estimate the geometric error between reference and detected building boundaries. For definitions and formula see (Awrangjeb et al., 2010).

4.1 Data Sets

The test data sets cover two urban areas in Queensland, Australia: Aitkenvale (AV) and Hervey Bay (HB). The AV area comprises two scenes. The first scene (AV1) covers an area of $108 \text{ m} \times 80$

m and contains 63 buildings. The second (AV2) covers an area of $66 \text{ m} \times 52 \text{ m}$ and contains 5 buildings. The HB data set has one scene and covers $108 \text{ m} \times 104 \text{ m}$ and contains 25 buildings. All three data sets contain mostly residential buildings and they can be characterised as urban with medium housing density and moderate tree coverage that partially covers buildings. AV2 comprises more vegetation and small buildings than AV1 and HB. In terms of topography, AV is flat while HB is moderately hilly.

For both AV and HB data sets, LIDAR coverage comprises first-pulse returns with a point density of 35 and 13 points/ m^2 , respectively, and the image data comprise multispectral orthoimagery with resolutions of 0.05 m and 0.2 m, respectively. Bare-earth DTMs of 1 m horizontal resolution cover both areas.

Two dimensional reference data sets were created by monoscopic image measurement using the Barista software (Barista, 2011). All visible buildings were digitised as polygons irrespective of their size. The reference data included garden sheds, garages, etc. These were sometimes as small as 5 m^2 in area.

4.2 Parameter Setting

The proposed building detection uses a number of parameters. Majority of them have been adopted from the existing research works. For example, the height threshold $h_c = 1 \text{ m}$, the mask resolution $r_b = 0.25 \text{ m}$ and the smallest building size 10 m^2 are from (Awrangjeb and Fraser, 2014); the minimum building width $W_m = 3 \text{ m}$ is from (Rottensteiner et al., 2005); and maximum point-to-point distance d_m is from (Awrangjeb et al., 2010). The gradient threshold $T_p = 0.15 \text{ m}$ is similar to the normal distance from point to plane (Awrangjeb and Fraser, 2014). Since, there is a height error in the supplied LIDAR data, all points at the same height usually do not have the same recorded height value. This error is sometimes in the range of $\pm 0.30 \text{ m}$. Thus, the use of $T_p = 0.15 \text{ m}$ is even lower than the estimated error. The angle threshold between successive bins of lines $\theta_d = \frac{\pi}{32} \text{ rad}$ is also smaller than existing threshold ($\frac{\pi}{16} \text{ rad}$) used for the principal directions of buildings (Awrangjeb, 2016). A small value of θ_d will allow to find more dominant directions of lines. However, in a given scene if it is observed that the number of dominant directions could be low, $\theta_d = \frac{\pi}{16} \text{ rad}$ can be used. The use of 1 m^2

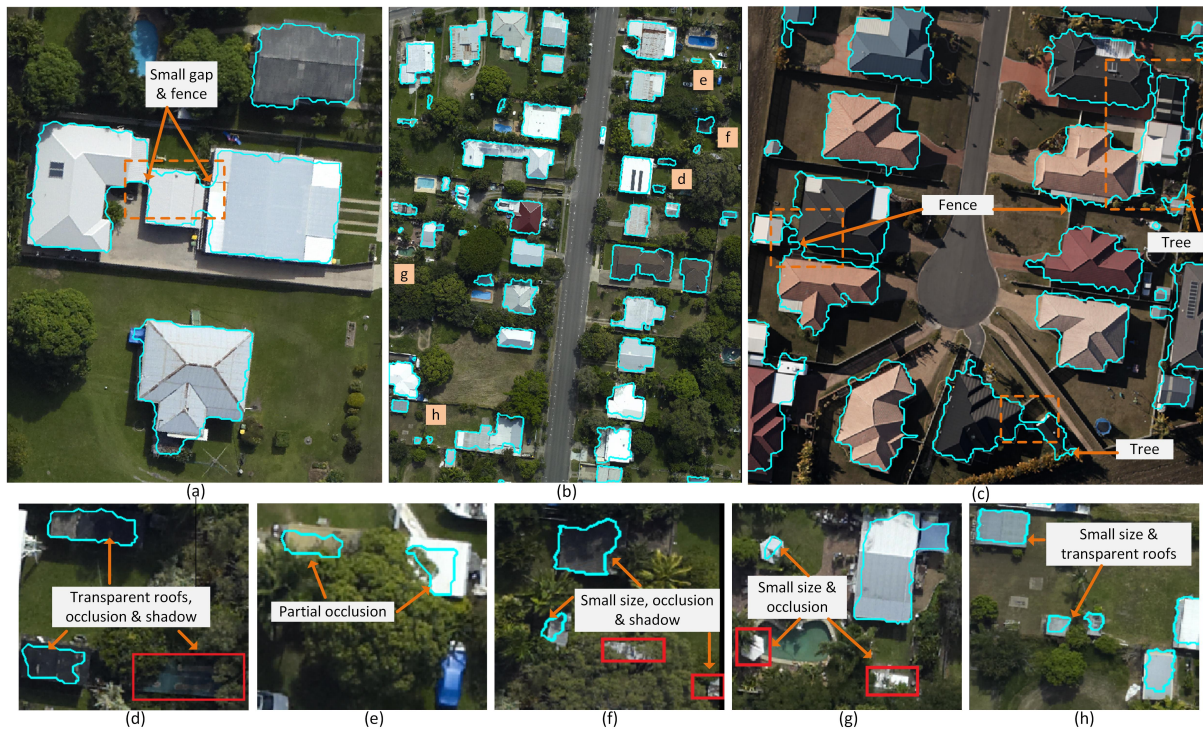


Figure 6. Building detection in (a-b) Aitkenvale (AV1 & AV2) and (c) Hervey Bay (HB) data sets, (d-h) some difficult areas in AV2 data set [marked with letters in (b)]. Cyan coloured polygons indicate detected buildings and red coloured polygons show missing buildings.

threshold as the smallest plane or building area and morphological filter size is also justified, since for the given the point density buildings and sheds smaller than 5 m² are hard to detect.

4.3 Experimental Results

Figures 6(a-c) show building detection results visually. As can be seen, all buildings in AV1 and HB data sets were correctly detected. However, many of the buildings were found merged (over-segmentation) with neighbouring buildings due to very little or no gaps between buildings in these two scenes. Some such errors are shown within orange coloured dotted rectangles in Figures 6(a) and (c). This merging error was observed less in number in AV2 data set where most of the buildings are well separated. But because of the occlusion and shadow problems, caused by dense vegetation, some small buildings could not be detected in AV2 data set.

Table 1 shows the object-based evaluation results on the test data sets. As can be seen, completeness, correctness and quality values for both AV1 and HB data sets are at the maximum (100%). However, the detection cross-lap rate (over-segmentation error) was high in these data set. In contrast, over-segmentation error was lower in AV2 data set, which is a quite complex data set and where some small buildings were missed due to shadow and severe occlusion. In this data set, completeness, correctness and quality values went up for buildings larger than 10 m² in area.

Figures 6(d-h) describe some difficult areas from AV2 data set. These areas are marked by letters in Figure 6(b). Some buildings, indicated within red coloured polygons, could not be detected mainly because of small in size, high occlusion and shadow. Small buildings do not have enough points reflected from their roofs to capture them into the proposed gradient-based building mask. Occlusion by trees offers high gradient measurements and,

Table 1. Object-based evaluation results in AV1 - first Aitkenvale, AV2 - second Aitkenvale and HB - Hervey Bay scenes. (C_m = completeness, C_r = correctness, Q_l = quality, C_{rd} = detection cross-lap rate and C_{rr} = reference cross-lap rate in percentage. $C_{m,10}$, $C_{r,10}$ and $Q_{l,10}$ are completeness, correctness and quality values for buildings with at least 10 m² in area.)

Data set	C_m	C_r	Q_l	C_{rd}	C_{rr}	$C_{m,10}$	$C_{r,10}$	$Q_{l,10}$
AV1	100	100	100	33.3	0	100	100	100
AV2	82.8	98	81.4	3.8	3.2	92.3	98	90.6
HB	100	100	100	22.2	9.1	100	100	100
Average	94.3	99.3	93.8	19.8	4.1	97.4	99.3	96.9

Table 2. Pixel-based and geometric evaluation results in AV1 - first Aitkenvale, AV2 - second Aitkenvale and HB - Hervey Bay scenes. (C_{mp} = completeness, C_{rp} = correctness, Q_{lp} = quality, B_f = branching factor and M_f = miss factor in percentage. RMSE = root-mean-square-error in metre.)

Data set	C_{mp}	C_{rp}	Q_{lp}	B_f	M_f	RMSE
AV1	92.8	97.2	90.4	2.9	7.7	0.63
AV2	84.2	96.8	81.9	3.3	18.7	0.89
HB	95	79.7	76.4	25.5	5.3	0.58
Average	90.7	91.2	82.9	10.6	10.6	0.70

thus, occluded buildings are hard to detect. Shadow increases the average NDVI value within a building area. As a result, candidate buildings exhibiting high NDVI values were removed. Moreover, some of the detected and missing buildings in Figure 6(g) are not only small in size and partially occluded but also of hexagonal sheds with triangular roof planes. The proposed building mask could not capture these small triangular planes well because of high gradient values between small triangular planes and occlud-

ing trees.

Table 2 shows the pixel-based and geometric evaluation results on the test data sets. Pixel-based completeness, correctness and quality values are more than 90% for AV1 data set, and more than 80% for AV2 data set. Nevertheless, the correctness value (hence, its quality value) for HB data set went down below 80% due to high over-segmentation errors, which are shown within an orange coloured dotted rectangles in Figure 6(c). Some detection polygons cover not only more than one actual buildings, but also surrounding fence and trees. This is also evident from a high branching factor (25.5%) for HB data set. Since, some small buildings were missed and many buildings could not be completely extracted due to occlusion in AV2 data set, its miss factor is higher (18.7%) than that of AV1 (7.7%) and HB (5.3%) data sets.

In terms of geometric error, *RMSE* value is higher for AV2 data set than for AV1 and HB data sets. Due to partial occlusions in many of the detected buildings in AV2 data set (see detected buildings shown in Figures 6(d)-(h)) they could not be completely extracted. These partially extracted buildings also contributed to the high miss factor for AV2 data set.

4.4 Comparative Results

Earlier the three test data sets were used for performance evaluation in (Awrangjeb and Fraser, 2014). In terms of object-based evaluation, (Awrangjeb and Fraser, 2014) performed the same as the proposed method in AV1 and HB data sets. However, in AV2 data set the proposed method outperformed (Awrangjeb and Fraser, 2014). While the completeness and quality values were below 70% by (Awrangjeb and Fraser, 2014), they are above 80% by the proposed method. This is because the proposed method is able to extract many transparent buildings exist in AV2 data set. The pixel-based correctness was also higher by the proposed method.

Despite the high object-based success of the proposed method, it has following limitations. Firstly, its pixel-based accuracy in terms of completeness and quality is slightly lower than that in (Awrangjeb and Fraser, 2014). Secondly, it does not work well with low density point cloud data. The proposed gradient-based mask cannot capture the small roof segments due to a low number of points reflected from roofs. It was tested with point density less than 10 points/m² and found ineffective. However, since high density point cloud data are becoming cheaper and more widely available now-a-days, the application of the proposed method to capture transparent and small buildings would offer a competitive alternative, specially where a high object-based performance is required, for example, the detection of buildings in remote country side (e.g., hilly and/or highly vegetated areas) for disaster (e.g., bushfire) management.

5. CONCLUSION

This paper has proposed a new building detection technique from the high density point cloud data and multispectral imagery. The newly proposed building mask is based on the gradient calculation on the height grid obtained at the dominant directions of non-ground image lines. Due to application of the gradient function on the height grid, any vegetation with random 3D surface is mostly removed. Thus, buildings become easily identifiable. As the proposed mask uses the non-ground points, it can capture the buildings with transparent roof materials. As a result,

the proposed mask integrates the benefits of two existing masks: the nDSM and the primary mask. Some false candidate buildings, specially found on planar surfaces of some vegetation, are removed with the application of NDVI from the multispectral imagery. The experimental results have shown its higher object-based performance than that of an existing building detection technique (Awrangjeb and Fraser, 2014). Moreover, it is capable of detecting small garden sheds, which are sometimes as small as 5 m² in area.

However, the proposed technique has some drawbacks. Its pixel-based accuracy is slightly lower than the existing technique and it does not work with low density point cloud data. The future work includes investigation of more effective building mask that would be more effective, specially with increased pixel-based accuracy and reduced over-segmentation error.

ACKNOWLEDGMENT

The authors would like to thank Ergon Energy (www.ergon.com.au) for providing the Australian data sets. Dr. Awrangjeb is a recipient of the Discovery Early Career Researcher Award by the Australian Research Council (project number DE120101778).

REFERENCES

- Awrangjeb, M., 2016. Using point cloud data to identify, trace, and regularize the outlines of buildings. *International Journal of Remote Sensing* 37(3), pp. 551–579.
- Awrangjeb, M. and Fraser, C. S., 2012. An automatic and threshold-free performance evaluation system for building extraction techniques from airborne LIDAR data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7(10), pp. 4184–4198.
- Awrangjeb, M. and Fraser, C. S., 2014. Automatic segmentation of raw LIDAR data for extraction of building roofs. *Remote Sensing* 6(5), pp. 3716–3751.
- 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* 65(5), pp. 457–467.
- Awrangjeb, M., Zhang, C. and Fraser, C. S., 2012. Building detection in complex scenes thorough effective separation of buildings from trees. *Photogrammetric Engineering & Remote Sensing* 78(7), pp. 729–745.
- Barista, 2011. The barista software.
- Habib, A. F., Zhai, R. and Changjae, K., 2010. Generation of complex polyhedral building models by integrating stereo-aerial imagery and lidar data. *Photogrammetric engineering & remote sensing* 76(5), pp. 609–623.
- Lee, D., Lee, K. and Lee, S., 2008. Fusion of lidar and imagery for reliable building extraction. *Photogrammetric Engineering & Remote Sensing* 74(2), pp. 215–226.
- 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.
- 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.