AN IMPROVED BUILDING DETECTION TECHNIQUE FOR COMPLEX SCENES

Mohammad Awrangjeb, Chunsun Zhang, Clive S. Fraser

Cooperative Research Centre for Spatial Information, Department of Infrastructure Engineering
University of Melbourne, Vic 3010, Australia

Phone: +61 3 8344 9182, Fax: +61 3 9654 6515, Email: {mawr, chunsunz, c.fraser}@unimelb.edu.au

ABSTRACT

The success of automatic building detection techniques lies in the effective separation of buildings from trees. This paper presents an improved automatic building detection technique that achieves more effective separation of buildings from trees. Firstly, it uses cues such as height to remove objects of low height such as bushes, and width to exclude trees with small horizontal coverage. The height threshold is also used to generate a ground mask where buildings are found to be more separable than in a so-called normalized DSM (digital surface model). Secondly, image entropy and colour information are jointly applied to remove easily distinguishable trees. Finally, an innovative rule-based procedure is employed using the edge orientation histogram from the imagery to eliminate false positive candidates. While tested on a number of scenes from four different test areas, the improved algorithm performed well even in complex scenes which are hilly and densely vegetated.

Index Terms— Automatic, building, detection, LIDAR, orthoimage, trees

1. INTRODUCTION

Building detection from remotely sensed data has a number of practical applications including city planning and disaster management. Therefore, a large number of building detection techniques have been reported over the last few decades. Since photogrammetric imagery and LIDAR (LIght Detection And Ranging) data have their own merits and demerits, the recent trend is to integrate data from both of these sources as a means of advancing building detection by compensating the shortcomings of one with the advantages of the other.

Nevertheless, the success of automatic building detection is still largely impeded by scene complexity, incomplete cue extraction and sensor dependency of data [1]. Vegetation, and especially trees, can be the prime cause of scene complexity and incomplete cue extraction. The situation becomes complex in hilly and densely vegetated areas where only a few buildings are present, these being surrounded by trees. Important building cues can be completely or partially missed due to occlusions and shadowing from trees. Therefore, many

existing building detection techniques that depend largely on colour information exhibit poor detection performance.

Application of a recently developed building detection algorithm [2] has shown it to be capable of detecting buildings in cases where cues are only partially extracted. For example, if a section of the side of a roof (at least 3m long) is correctly detected, the algorithm can also detect all or part of the entire building. However, this detector does not necessarily work well in complex scenes when buildings are surrounded by dense vegetation and when they have the same colour as trees, or where trees are other than green.

This paper, which is a condensed version of [3] but contains additional experimental results on a new test site, presents an improved detection algorithm that uses both LIDAR and imagery. In addition to exploiting height, width and colour information, it uses different texture information in order to differentiate between buildings and trees. Firstly, objects below a given threshold above the ground, such as bushes, cars and carports, are removed from the raw LIDAR data. This cue is also used to generate a ground mask where trees are found to be more separable than in the well known normalized digital surface model (nDSM). Secondly, image entropy and colour information are employed together to remove the trees that are easily distinguishable. Finally, false positive detections are eliminated using an innovative rule-based procedure based on the edge orientation histogram. The improved detector has been tested on a number of scenes covering four different test areas.

2. CUES TO IDENTIFY TREES

A height threshold (2.5m above ground level) is often used to remove low vegetation and other objects of limited height, such as cars and street furniture [2]. The height difference between first and last pulse LIDAR ranging data has also been used [4]. Some algorithms use width and area attributes to remove trees. If the width or area of a detected object is smaller than a threshold, then it is removed as a tree [2].

Texture information has also been employed as an important cue to identify trees. When objects have similar spectral responses, the grey level co-occurrence matrix (GLCM) can be estimated from the image to quantify the co-occurrence

Fig. 1. Proposed improved building detection technique.

probability [5]. Some GLCM indices, eg mean, standard deviation, entropy and homogeneity, have been applied to both height and image data in order to classify buildings and trees [6]. Rottensteiner et al. [7] applied a polymorphic feature extraction algorithm to the first derivatives of the DSM in order to estimate the surface roughness. Zhang et al. [8] applied a plane-fitting technique to non-ground LIDAR points to identify trees. While a satisfactory fitting is expected for building planes, there will likely be large gross errors for trees.

Colour information is frequently used to eliminate trees. While a high NDVI (normalised difference vegetation index estimated using multispectral images) value represents a vegetation pixel, a low NDVI value indicates a non-vegetation pixel. However, this cue has been found unreliable even in normal scenes where trees and buildings have distinct colours [7]. Vu et al. [9] applied the K-means clustering algorithm on multispectral images to obtain spectral indices for clusters like trees, water and buildings. Shorter and Kasparis [10] used colour invariants. A number of other cues generated from colour image and height data can be found in [6].

3. PROPOSED DETECTOR

The proposed detector, which is based on a previous detector [2], employs a combination of height, width, colour and texture information with the aim of more comprehensively separating buildings from trees. Although cues other than texture were used in [2], the improved formulation makes use of additional texture cues such as entropy and the edge orientation histogram at three stages of the process, as shown in Fig. 1. Different steps of the detection algorithm have been presented in [2]. This paper focuses on how texture, dimensional and colour information can be applied jointly in order to better distinguish between buildings and trees.¹

¹ All parameter values presented in this section were set empirically and found insensitive, not only to their value changes, but also while applied to

Fig. 2. (a) Images of two test scenes, (b) corresponding LIDAR data (shown in gray-scale), (c) ground masks and (d) normalised DSMs (shown in binary). Black pixels in nDSMs indicate non-zero heights greater than the height threshold.

3.1. Height

A height threshold $T_h = H_g + 2.5\text{m}$, where H_g represents the ground or DEM (digital elevation model) height, is applied to the raw LIDAR data. This threshold removes objects of low height (shrubbery, road furniture, cars, etc.) and preserves non-ground points (trees and buildings).

The height threshold T_h is also used to generate a ground mask M_g , which has the same resolution as the image. The black areas in M_g indicate *void areas* where there are no laser returns below T_h (ground areas covered by buildings and trees). This mask is technically different from the well known nDSM where non-zero heights indicate the elevated objects above T_h . Therefore, as shown in Fig. 2, buildings and trees are found to appear thinner in M_g than in the nDSM. For the first scene in Fig. 2, most of the trees around the building at the lower left of the scene (shown within a circle) will be clearly separable in M_g , while in the nDSM they are almost connected to the building. For the second scene in Fig. 2, each of the buildings is strongly connected to the neighbouring vegetation in the nDSM, while they are clearly separable in M_g . Consequently, unlike the existing building detectors that use the nDSM for building detection, the use of M_g to obtain building candidates by the proposed detector helps to better separate trees from buildings.

3.2. Width, NDVI and Entropy

The black areas in M_g are either buildings, trees or other elevated objects. Line segments around these black shapes in M_g are formed and segments shorter than the minimum building width $L_{min} = 3\text{m}$ are removed. Trees having small horizontal area are thus removed.

The mean of the NDVI value is then applied on the remaining line segments, as described in [2], to eliminate trees

different data sets discussed in Section 4.

Fig. 3. Detection of green buildings: (a) the NDVI information alone missed green buildings whereas (b) combined NDVI and entropy information detects green buildings. ‘Blue’ lines are accepted, ‘red’ represents rejected.

Fig. 4. A complex scene: (a) ground mask, (b) detected candidate buildings with a large number of false detections and (c) detected final buildings after removing false positives.

covering a large horizontal area. However, the NDVI has been found to be an unreliable cue [7] and it cannot differentiate between trees and green buildings. Fig. 3(a) shows an example where a green building B_1 cannot be detected at all since all lines around it are rejected using the NDVI. However, green building B_2 can be partially detected because it has a white coloured roof section. In some areas there may be non-green buildings having the same colour as trees, especially when leaves change colour in different seasons. In such cases, the removal of trees based on the NDVI will result in many buildings also being removed. Detection of these same buildings will likely also lead to detection of trees.

If the mean NDVI is above the NDVI threshold at any side of a line segment, a further test is performed before removing this line segment as a tree-edge. This test checks whether the average entropy is more than the entropy threshold $T_{ent} = 30\%$. If the test holds, the line segment is removed as a tree edge, otherwise it is selected as a building edge. Fig. 3(b) shows that the green buildings B_1 and B_2 can be fully detected using this approach. In addition, some of the trees subject to shadowing and self-occlusion are also detected.

3.3. Edge Orientation Histogram

In areas with dense vegetation, the black shapes of buildings and nearby trees in M_g are not separable and consequently a

building may be connected with another building a few metres away (see Fig. 4). Therefore, a large number of false buildings may be present in the candidate building set, as shown in Fig. 4(b). A procedure utilising the edge orientation histogram from the orthoimage is therefore developed in order to remove these false positives.

A gradient histogram is formed using the edge points within each candidate building rectangle. Edges are first extracted from the orthophoto using an edge detector and short edges (less than 3m in length) are removed. Each edge is then smoothed and the gradient (tangent angle) is calculated on each point using the first order derivatives. The gradient will be in the range $[-90^\circ, +90^\circ]$. A histogram with a successive bin distance of $D_{bin} = 5^\circ$ is formed using the gradient values of all edge points lying inside the candidate rectangle.

Rectangles containing the whole or major part of a building should have one or more significant peaks in the histogram, since edges detected on building roofs are formed from straight line segments. All points on an apparent straight line segment will have a similar gradient value and hence will be assigned to the same histogram bin, resulting in a significant peak. A significant peak means the corresponding bin height is well above the mean bin height of the histogram. Since edge points whose gradient falls into the first (at -90° to -85°) and last (at 85° to 90°) bins have almost the same orientation, located peaks in these two bins are added to form a single peak.

Fig. 5 illustrates three gradient histogram functions and mean heights for candidate buildings B_1 , B_2 and B_3 in Fig. 4(b). Two bins at $\pm 90^\circ$ basically form one bin, because lines in these two bins are perpendicular to the x -axis and reside above and below this axis. Therefore, these can be a peak at either of these bins and their heights can be accumulated to form a single peak. Fig. 5(a) shows that B_1 has two significant peaks: 80 pixels at 0° and 117 ($55 + 62$) pixels at $\pm 90^\circ$, these being well above the mean height of 28.6 pixels. The two significant peaks separated by 90° strongly suggest that this is a building. From Fig. 5(b) it can be seen that B_2 has one significant peak at $\pm 90^\circ$ but a number of insignificant peaks. This points to B_2 being partly building but mostly vegetation, which is also supported by the high mean height value. With the absence of any significant peak, but a number of insignificant peaks close to the mean height, Fig. 5(c) indicates that B_3 is comprised of vegetation. Although there may be some significant peaks in heavily vegetated areas, a high average height of bins between two significant peaks can be expected. Note that the orthophoto resolution in this case was 10cm, so a bin height of 80 pixels indicates a total length of 8m from the contributing edges.

The observations above support the theoretical inferences. In practice, however, detected vegetation clusters can show the edge characteristics of a building, and a small building having a flat roof may not have enough edges to show the required peak properties. As a result, some true buildings can

Fig. 5. Gradient histogram functions and means for rectangles (a) B_1 , (b) B_2 and (c) B_3 in Fig. 4(b): x -axis is in degrees and y -axis is in pixels (bin heights).

be missed, while some false buildings may be detected. A number of precautions can be formulated in order to minimize the occurrence of false detections.

Two types of histograms are formed using edges within each detected rectangle. In the first type, one histogram considers all the edges collectively, and in the second type histograms for individual edges whose length is at least L_{min} are formed. Let the collective histogram be symbolized as H_{col} , with an individual histogram being indicated by H_{ind} . Tests on H_{col} and H_{ind} can be carried out to identify true buildings and remove trees. If a detected rectangle passes at least one of the following tests it is selected as a building, otherwise it is removed as vegetation.

1. *Test 1:* H_{col} has at least two peaks with heights of at least $3L_{min}$ and the average height of bins between those peaks is less than $2L_{min}$. This test ensures the selection of a large building, where at least two of its long perpendicular sides are detected. It also removes vegetation where the average height of bins between peaks is high.
2. *Test 2:* The highest bin in H_{col} is at least $3L_{min}$ in height and the aggregated height of all bins in H_{col} is at most 90m. This test ensures the selection of a large building where at least one of its long sides is detected. It also removes vegetation where the aggregated height

of all bins is high.

3. *Test 3:* H_{col} has at least two peaks with heights of at least $2L_{min}$, and the highest bin to mean height ratio R_{Mm1} is at least 3. This test ensures the selection of a medium size building, where at least two of its perpendicular sides are detected. It also removes vegetation where the highest bin to mean height ratio is low.
4. *Test 4:* The highest bin in H_{col} has a height of at least L_{min} and the highest bin to mean height ratio R_{Mm2} is at least 4. This test ensures the selection of a small or medium size building where at least one of its sides is at least partially detected. It also removes small to moderate sized vegetation areas where the highest bin to mean height ratio is low.
5. *Test 5:* The highest bin in H_{ind} has a height of at least L_{min} and the aggregated height of all bins in H_{col} is at most 90m. This test ensures the selection of buildings which are occluded on at most three sides.
6. *Test 6:* The ratio R_{aTp} of the detected rectangular area to the number of texture pixels (N_{Tp} , the aggregated height of all bins in H_{col}) is at least 45. This test ensures the selection of all buildings which are at least partially detected but the roof sides are missed.

The application of these tests on the complex scene in Fig. 4(b) produces the result shown in Fig. 4(c).

4. PERFORMANCE STUDY

The evaluation system involved in the performance study conducted makes one-to-one correspondences using nearest centre distances between detected and reference buildings. Some 15 evaluation indices in three categories, namely object-based, pixel-based and geometric, have been employed. Whereas pixel-based evaluation considers performance based on the number of pixels in an object in the image, object-based evaluation takes into account the number of objects in the image. The root mean square positional discrepancy value (RMSE) is employed to quantify the geometric accuracy. The detailed procedure of the threshold-free evaluation system and the evaluation indices can be found in [11].

The test data sets employed cover four suburban areas in Australia, Aitkenvale, QLD; Fairfield, NSW; Moonee Ponds, Victoria and Knox, Victoria. The Aitkenvale data set covers an area of 291m $\times$ 234m and contains 87 buildings, Fairfield covers 588m $\times$ 417m and contains 370 buildings, Moonee Ponds covers 447m $\times$ 447m and has 250 buildings and Knox covers 400m $\times$ 400m and contains 130 buildings. Fairfield contains many large industrial buildings and in Moonee Ponds there were some green buildings. Knox can be characterized as outer suburban with lower housing density and extensive

tree coverage that partially covers buildings. In terms of topography, Aitkenvale, Fairfield and Moonee Ponds are relatively flat while Knox is quite hilly.

LIDAR coverage comprised a point spacing of 0.17m for Aitkenvale, 0.5m for Fairfield, and 1m for both Moonee Ponds and Knox. For Aitkenvale, Fairfield and Knox, RGB colour orthoimagery was available, with resolutions of 0.05m, 0.15m and 0.1m, respectively. Moonee Ponds image data comprised RGBI colour orthoimagery with a resolution of 0.1m. Bare-earth DEMs of 1m horizontal resolution covered all four areas.

Reference data sets B_r were created by monoscopic image measurement using the Barista software.² All rectangular structures above the height threshold T_h were digitized. Therefore, B_r included buildings, garden sheds and garages which were sometimes as small as 10m² in area.

Tables 1 to 3 show results of the object-based, pixel-based and geometric accuracy evaluations of the improved building detection algorithm in the four test areas. A visual illustration of sample building detection results is shown in Fig. 6. The reason for showing the visual results using Knox is that scenes from Knox are more complex (height variation and tree colours) than those from Fairfield and Moonee Ponds.

The improved algorithm produced moderately better performance than [2] in all three evaluation categories within Aitkenvale, Fairfield and Moonee Ponds. The better performance was mainly due to proper detection of large industrial buildings in Fairfield, detection of some green buildings in Moonee Ponds, and elimination of trees in Aitkenvale, Fairfield and Moonee Ponds.

In Knox, the proposed improved algorithm exhibited significantly better performance over Awrangjeb et al. [2], due to two main reasons. Firstly, the improved algorithm better accommodated the dense tree cover and randomly oriented buildings that characterized the Knox data. Aitkenvale, Fairfield and Moonee Ponds on the other hand are low in vegetation cover and buildings are generally well separated from each other. Secondly, the improved algorithm showed its merits in better handling varying topography. Knox is a hilly area (maximum height $H_M = 270$ m and minimum height $H_m = 110$ m), whereas Aitkenvale ($H_M = 69$ m and $H_m = 60$ m), Fairfield ($H_M = 23$ m and $H_m = 1$ m) and Moonee Ponds ($H_M = 43$ m and $H_m = 23$ m) are almost flat.

Awrangjeb et al. [2] detected a large number of false buildings in Knox, as illustrated in Fig. 6(a). So, in object-based evaluation, 56% quality was observed with 77% completeness and 67% correctness. The reference cross-lap rate was above 85%, with 39% detection overlap rate. In pixel-based evaluation, 27% quality was found with 44% completeness and 42% correctness. The area omission error was more than 50% and both branching and miss factors were above 120%. The geometric accuracy was no better than 3.3m.

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}
Aitkenvale	97.7	98.8	96.6	5.5	13.2	14.3	27.6
Fairfield	95.1	95.4	92.2	2.7	8.6	3.5	9.7
MPonds	94.5	95.3	89.2	6.2	13.1	7.3	17.5
Knox	93.2	87.2	82.0	9.3	12.8	23.3	61.6
Average	95.1	94.2	90	5.9	11.9	12.1	29.1

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
Aitkenvale	94.5	98.9	97.4	5.7	3.5	4.5	5.8
Fairfield	83.2	84.5	72.4	15.3	12.5	13.5	20.3
MPonds	87.2	85.4	75.3	12.7	13.2	16.7	17.3
Knox	74.1	77.3	60.9	26.4	21.0	29.4	34.9
Average	84.8	86.5	76.5	15	12.6	16	19.6

Table 3. Geometric accuracy.

Scenes	metres	pixels
Aitkenvale	1.1	22.0
Fairfield	2.4	16.0
Moonee Ponds	1.6	16.0
Knox	2.0	20.0
Average	1.8	18.5

In contrast, as shown for Knox in Fig. 6(b), the improved detector removed a large number of false buildings using its orientation histogram. In object-based evaluation, when compared to [2], the quality increased to 82%, a 26% rise. The detection overlap rate decreased to 13% and the reference cross-lap rate reduced to 62%. In pixel-based evaluation, again when compared to [2], the quality went up to 61%, a 43% growth, while the branching factor declined dramatically to 29% and the miss factor was also considerably improved to 35%. Geometric accuracy improved to 20 pixels, or by approximately 50%.

In object-based evaluation, the improved algorithm offered on average across the four data sets a more than 10% increase in completeness and correctness and a 15% increase in quality. Multiple detection and detection overlap rates were also low. In pixel-based evaluation, there was also a reasonable rise in completeness (4%), correctness (10%) and quality (7%). Area omission and commission errors were less than those obtained with the original algorithm. In addition, there was a 5 pixel improvement in geometric accuracy.

²The Barista Software, www.baristasoftware.com.au, May 2011.

Fig. 6. Building detection on a complex scene from Knox by (a) Awrangjeb et al. [2] and (b) improved algorithm.

5. CONCLUSIONS

We have presented an improved automatic building detection technique that exhibits better performance in separating buildings from trees. In addition to employing height and width thresholds and colour information, it uses texture information from both LIDAR and colour orthoimagery. The joint application of measures of entropy and NDVI helps in the removal of vegetation by making trees more easily distinguishable. The newly proposed innovative rule-based procedure, based on the edge orientation histogram from the image edges, assists in eliminating a large number of false positive candidates in complex scenes. The experimental results reported showed that while the improved algorithm offered moderately enhanced performance in Aitkenvale, Fairfield and Moonee Ponds, it yielded a very significant improvement in performance in Knox across all three evaluation categories.

6. REFERENCES

- [1] 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*, vol. 62, no. 1, pp. 43–63, 2007.
- [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*, vol. 65, no. 5, pp. 457–467, 2010.
- [3] M. Awrangjeb, C. Zhang, and C. S. Fraser, "Building detection in complex scenes thorough effective separation of buildings from trees," *Photogrammetric Engineering and Remote Sensing*, vol. 78, no. 7, 2012.
- [4] 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*, vol. 37, no. part B7, pp. 1183–1188, 2008.
- [5] L. Chen, T. Teo, C. Hsieh, and J. Rau, "Reconstruction of building models with curvilinear boundaries from laser scanner and aerial imagery," *Lecture Notes in Computer Science*, vol. 4319, pp. 24–33, 2006.
- [6] M. Salah, J. Trinder, and A. Shaker, "Evaluation of the self-organizing map classifier for building detection from lidar data and multispectral aerial images," *Journal of Spatial Science*, vol. 54, no. 2, pp. 1–20, 2009.
- [7] 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 J of Photogrammetry and Remote Sensing*, vol. 62, no. 2, pp. 135–149, 2007.
- [8] K. Zhang, J. Yan, and S. C. Chen, "Automatic construction of building footprints from airborne lidar data," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 44, no. 9, pp. 2523–2533, 2006.
- [9] 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*, vol. 11, no. 4, pp. 281–289, 2009.
- [10] N. Shorter and T. Kasparis, "Automatic vegetation identification and building detection from a single nadir aerial image," *Remote Sensing*, vol. 1, no. 4, pp. 731–757, 2009.
- [11] M. Awrangjeb, M. Ravanbakhsh, and C. S. Fraser, "Building detection from multispectral imagery and lidar data employing a threshold-free evaluation system," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 38, no. part 3A, pp. 49–55, 2010.