EFFECTIVE BUILDING DETECTION IN COMPLEX SCENES

Mohammad Awrangjeb

Gippsland School of Information Technology
Monash University, Vic 3842, Australia
E-mail: mohammad.awrangjeb@monash.edu

Clive S. Fraser

CRC for Spatial Information
University of Melbourne, Vic 3010, Australia
E-mail: c.fraser@unimelb.edu.au

ABSTRACT

Separation of buildings from trees is a major challenge in automatic building detection. In residential and hilly areas, buildings are often surrounded by dense vegetation. This paper presents a three-step method for effective separation of buildings from trees. Firstly, height and width thresholds are applied to LIDAR data for removing small bushes and trees with small horizontal coverage, respectively. The generation of the building mask, where each black region indicates a void area from which there are no laser returns below the height threshold, also helps in separation of buildings from the nearby trees. Then image entropy and colour information are applied together to remove trees exhibiting high texture. Finally, an innovative rule-based procedure is employed using the edge orientation histogram from the imagery to eliminate the remaining trees. Experimental results show that the algorithm offers high building detection rate in complex scenes which are hilly and densely vegetated.

***Index Terms*—** Automatic, building detection, LIDAR, orthoimage, separation, 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 [1]. Nevertheless, the success of automatic building detection is still largely impeded by scene complexity, incomplete cue extraction and sensor dependency of data [2]. 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.

A recently developed building detection algorithm [1] has shown that it is 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, especially in hilly areas, when buildings are surrounded by dense vegetation.

This paper presents an improved detection algorithm that uses both LIDAR data and aerial imagery. In addition to exploiting height, width and colour information, it uses different texture information in order to differentiate between buildings and trees. The improved detector has been tested on a number of scenes covering two different test areas in Australia and one in Germany from the ISPRS benchmark [3]. Note this paper is a condensed version of Awrangjeb et al. [4] and presents results and discussions on new data sets.

2. BUILDING DETECTION

The proposed detector, which is an improved version of a previous development [1], employs a combination of dimensional, colour and texture information with the aim of more comprehensively separating buildings from trees. Although cues other than texture have previously been used [1], the improved formulation makes use of additional texture cues such as entropy and the edge orientation histogram at three stages of the procedure (Fig. 1). Different steps of the detection algorithm have previously been presented in [1], and this paper focuses on how texture, dimensional and colour information can be applied jointly in order to better distinguish between buildings and trees, these processes being indicated within the dashed rectangle in Fig. 1.

A height threshold of $T_h = H_g + H_c$m, where H_g represents the ground or DEM (digital elevation model) height and H_c is a height constant, is first applied to the raw LIDAR data. This threshold removes objects of low height (shubbery, road furniture, cars, etc.) and preserves non-ground points (trees and buildings). The value of H_c depends on the data

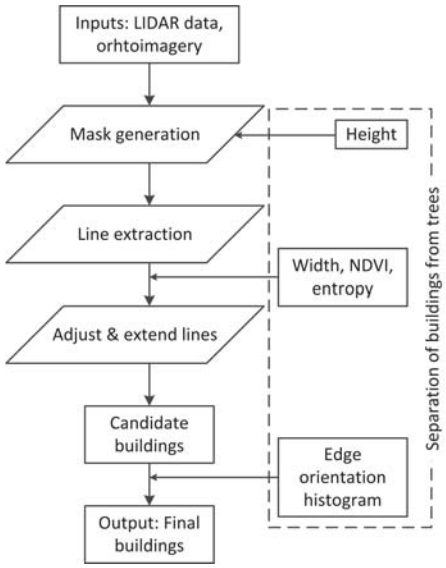

Fig. 1. Proposed building detection technique.

set. While in some area a large value, for instance 2.5m [5], may remove a lot of vegetation, in other area this value may remove some low height objects such as garages and sheds too. Thus it is wise to set a small value, for instance $H_c = 1$m to keep the small height objects and remove vegetation by other means discussed below. The height threshold T_h is also used to generate a ground mask M_g, whose resolution is set at 0.25m independent of the LIDAR density and input image resolution. 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). Line segments around these black shapes in M_g are formed and segments shorter than the minimum building width $L_{min} = 3$m are removed. Trees having small horizontal area are thus removed.

The mean of the NDVI (normalised difference vegetation index) value for multispectral imagery or pseudo-NDVI for RGB imagery is then applied on the remaining line segments, as described in [1]. 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. The remaining line segments are extended and adjusted and the candidate buildings are formed as rectangular shapes.

A gradient histogram is then formed using the edge points within each candidate building rectangle. Edges are first extracted from the imagery 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 at each point using first order derivatives. The gradient will be in the range $[-90°, +90°]$. A histogram with a successive bin distance of $D_{bin} = 5°$ 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. 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.

Fig. 2. Building detection by the proposed detector on Area 2 of the Eltham data set.

Fig. 3. Building detection by the proposed detector on Area 3 of the Vaihingen data set.

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.

3. TEST DATA SETS

The test data sets employed cover three suburban areas: two in Australia and one in Germany. The Australian sites are Harvey Bay, Queensland and Eltham, Victoria. There were two scenes from each of these areas. The Harvey Bay scenes were 239m $\times$ 214m and 156m $\times$ 228m in area, containing 37 and 64 buildings, respectively. The Eltham scenes were 241m $\times$ 262m and 222m $\times$ 390m in area and contained 51 and 89 buildings, respectively. Both areas mostly comprised residential buildings, but the Harvey Bay area also have some industrial buildings. While Harvey Bay is low in vegetation, Eltham can be characterized as outer suburban with lower housing density and extensive tree coverage that partially covers buildings, as indicated in Fig. 2. In terms of topography, Harvey Bay is relatively flat while Eltham is quite hilly. For both areas, LIDAR density was around 17 points/m^2 and RGB colour orthoimagery was available with resolutions of 0.10m for Harvey Bay and 0.05m for Eltham.

Table 1. Object-based evaluation results in percentages (C_m = completeness, C_r = correctness and Q_l = quality).

Scenes	C_m	C_r	Q_l
Harvey Bay Area 1	97.3	94.7	92.3
Harvey Bay Area 2	87.5	100	87.5
Eltham Area 1	90.2	78.0	71.9
Eltham Area 2	95.5	82.5	79.4
Vaihingen Area 1	81.1	87.9	72.9
Vaihingen Area 2	78.6	45.5	40.4
Vaihingen Area 3	67.9	95.0	65.5

The German site was the Vaihingen data set [6] adopted by the ISPRS benchmark [3]. The LIDAR point density varies between 4 to 6.7 points/m^2 and the RGB colour orthoimage has a resolution of 0.09m. There are three test areas in this data set. The number of buildings (larger than 2.5m^2) in these tree areas are 37, 14 and 56, respectively. Area 1 is characterized by dense development consisting of historic buildings having rather complex shapes, roads and trees. Area 2 is characterized by a few high-rising residential buildings that are surrounded by trees. Area 3 is a purely residential area with detached houses and many surrounding trees.

For the Australian data sets, the reference buildings were created by monoscopic image measurement using the Barista software [7]. All rectangular structures, recognizable as buildings and above the height threshold T_h, were digitized. The reference data included garden sheds, garages, etc. These were sometimes as small as 10m^2 in area. The building detection results were evaluated by means of an automatic evaluation system that makes one-to-one correspondences using nearest centre distances between detected and reference buildings [5].

For the German data set, the detection results were sent to the ISPRS commission III, Working Group 4 [1] and evaluated using a threshold-based evaluation system [8]. In the reference building sets, there were buildings which were as small as 2.5m^2.

4. RESULTS AND DISCUSSIONS

Tables 1 and 2 show the object- and pixel-based evaluation results by the proposed improved detector. The overall root

[1]http://www2.isprs.org/commissions/comm3/wg4/tests.html

Table 2. Pixel-based evaluation results in percentages (C_{mp} = completeness, C_{rp} = correctness and Q_{lp} = quality).

Scenes	C_{mp}	C_{rp}	Q_{lp}
Harvey Bay Area 1	73.5	66.3	53.5
Harvey Bay Area 2	59.9	67.6	46.5
Eltham Area 1	54.8	76.8	47.0
Eltham Area 2	63.2	69.1	49.3
Vaihingen Area 1	89.7	83.4	76.2
Vaihingen Area 2	87.9	66.1	60.6
Vaihingen Area 3	86.4	88.2	77.4

mean square error (RMSE) produced in the four Australian scenes was 3.4m. This large error was mainly due to the large registration error between the input LIDAR data and the orthoimagery. The Harvey Bay data set did not have metrically precise orthophotos and there was residual relief displacement observed in both data sets. The improved algorithm showed slightly better performance in the Harvey Bay data set, but significantly better performance in the Eltham data set (Fig. 2). The considerable improvement was due to better separation of buildings from trees by the improved detector. Note that the average object-based completeness, correctness and quality on four scenes by the original detector [1] were 93.9%, 70.5% and 65.7%. The average pixel-based values of these indices were 63.8%, 62.7% and 46.0% respectively. When these results are compared with those in Tables 1 and 2(Harvey Bay and Eltham results in first 4 rows), it is evident that the proposed detector offers significantly better performance.

The RMSE error for the three German scenes was 1.16m, 1.33 and 0.98m. Although the registration error between the LIDAR data and the orthoimage was low, there was moderate vegetation surrounding buildings in the first two scenes. The object-based evaluation in Area 2 was low because the buildings which are small in size (as small as 2.5m^2) could not be detected and some trees could not be removed.

While results from the Australian data sets are compared with those from the German data set, it is evident that the improved detector performed better in the Australian sites. The reason is that the LIDAR point density was higher in the Australian data sets and there were some small buildings in the German data set which could not be detected by the proposed detector.

5. CONCLUSIONS

The proposed improved automatic building detection technique has been shown to exhibit better performance in separating buildings from trees. This is due to the proposed innovative rule-based procedure, based on the edge orientation histogram from the image, which assists in eliminating a large number of false positive candidates in complex scenes.

Acknowledgment

Dr. Awrangjeb is the recipient of the Discovery Early Career Researcher Award by the Australian Research Council (project number DE120101778). The Vaihingen data set was provided by the German Society for Photogrammetry, Remote Sensing and Geoinformation (DGPF) [6]: www.ifp.uni-stuttgart.de/dgpf/DKEP-Allg.html. The Harvey Bay data set was from the Ergon Energy (www.ergon.com.au) and the Eltham data set was from the Department of Sustainability and Environment (www.dse.vic.gov.au) in Australia.

6. REFERENCES

[1] 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.

[2] 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.

[3] F. Rottensteiner, G. Sohn, J. Jung, M. Gerke, C. Baillard, S. Benitez, and U. Breitkopf, "The isprs benchmark on urban object classification and 3d building reconstruction," vol. I-3, pp. 293–298, 2012.

[4] M. Awrangjeb, C. Zhang, and C. S. Fraser, "Building detection in complex scenes thorough effective separation of buildings from trees," *Photogrammetric Engineering & Remote Sensing*, vol. 78, no. 7, 2012.

[5] 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.

[6] M. Cramer, "The dgpf test on digital aerial camera evaluation - overview and test design," *Photogrammetrie Fernerkundung Geoinformation*, vol. 2, pp. 73–82, 2010.

[7] Barista, "The barista software," 2011.

[8] M. Rutzinger, F. Rottensteiner, and N. Pfeifer, "A comparison of evaluation techniques for building extraction from airborne laser scanning," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 2, no. 1, pp. 11–20, 2009.