EFFECTIVE SEPERATION OF TREES AND BUILDINGS FOR AUTOMATED BUILDING DETECTION

Chunsun Zhang, Mohammad Awrangjeb, and Clive S. Fraser
Cooperative Research Centre for Spatial Information, The University of Melbourne
Level 5, 204 Lygon Street, Carlton Vic 3053, Australia
Phone: +61 3 8344 9182, Fax: +61 3 9349 5185
Email: {chunsunz, mawr, c.fraser}@unimelb.edu.au

KEY WORDS: Building Detection Lidar, Fusion, Texture, Classification

ABSTRACT: Effective separation of buildings from trees is a major challenge in automatic building detection from aerial imagery and Lidar data. In cases where an adopted building detection technique cannot distinguish between these two classes of objects, the presence of trees in the scene can increase the rates of both false positives and false negatives in the building detection process. This paper presents an automatic building detection technique which exhibits improved separation of buildings from trees. In addition to using traditional cues such as height, width and colour, the improved detector uses texture information from both Lidar and orthoimagery. Firstly, image entropy and colour information are jointly applied to remove easily distinguishable trees. Secondly, a rule-based procedure using the edge orientation histogram from the imagery is followed to eliminate false positive candidates. The improved detector has been tested on a number of scenes from three different test areas and it is shown that the algorithm performs well even in complex scenes with over 10% increase both in completeness and correctness.

1. INTRODUCTION

Automated building detection from remotely sensed data has been an active topic in photogrammetry and computer vision. Buildings are an indispensible component in a geospatial information system. In addition, some practical applications, including emergency response, homeland security and disaster (flood or bushfire) management require up-to-date building information, dictating the importance and necessity of timely acquisition of building information over large areas. As remote sensing imagery is the main source for spatial information generation, automated analysis of satellite and aerial images for building detection has been investigated (Mayer, 1999). Buildings were detected based on the optical reflectance of roof materials and/or with the knowledge of building shape information. The single image analysis techniques neglect the inherent 3D information. Therefore, multiple images were introduced with 3D information generated using photogrammetry techniques and later with Lidar data. Lidar technology provides dense accurate georeferenced 3D point clouds over reflected objects. The introduction of Lidar data has offered an attractive option for improving the level of automation in building detection process. Recent trend in building detection is to integrate Lidar data with imagery to benefit from the accurate 3D Lidar information and extensive 2D information such as high-resolution texture and color information in images for enhanced performance (Sohn and Dowman, 2007; Vu et al., 2009; Awrangjeb et al., 2010).

Despite significant efforts in research, full automated building detection still does not mature for commercialization. The success is largely impeded by scene complexity, incomplete cue extraction and sensor dependency of data (Sohn and Dowman, 2007). One of the challenges is the efficient separation of trees and buildings. Like buildings, trees are above ground objects in 3D data. Shadows and occlusions by tall trees nearby buildings cause inhomogeneous appearance of roof in remote sensing imagery. Tall trees also prevent Lidar strikes on roof, resulting incomplete 3D information of building roof. The situation becomes even more complex in hilly and densely vegetated areas. In this paper, we present a new approach to efficient separation of buildings and trees from aerial imagery and Lidar data. First, Lidar data is exploited to detect above ground objects, including buildings and trees. Then, texture analysis and edge orientation information are applied to distinguish buildings and trees. The remainder of this paper is organized in following. After this introduction, we briefly review the existing approaches to separation of buildings and trees. Then, our new method is explained in details. Tests were conducted using aerial imagery and Lidar data over various terrains and land covers. The results are presented together with evaluation using manually plotted reference data. The last section completes the paper with discussions and conclusions.

2. BRIEF LITERATURE REVIEW

Existing building detection algorithms make use of different cues with a view to separating buildings from trees. While cues related to colour are only available with multispectral images, cues related to width, height and area can be derived from Lidar or images. 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 (Rottensteiner et al., 2007; Awrangjeb et al., 2010). Trees taller than the building roof cannot be removed via this height threshold. Dash et al. (2004) used the height variation along the periphery of objects present in the data to distinguish trees from buildings. (Rottensteiner et al., 2007) and Khoshelham et al. (2008) used height difference values between first and last pulse Lidar data for the same purpose, since it can be anticipated that the differences will be large for trees but negligible for buildings. However, a first pulse is not always reflected from the upper branches of a tree and a last pulse may sometimes be a reflection from a tree trunk or branches (Maas, 2001).

Approaches based on segment classification of Lidar point clouds have been developed and segment attributes were exploited for differentiation of buildings and trees. Segments can be generated by plane-fitting techniques on the non-ground Lidar points (Zhang et al., 2006), or region growing methods based on seed points detected with 3D Hough transformation (Vosselman et al., 2004). Sampath and Shan (2010) reported a segmentation approach employed eigenanalysis to yield surface normal and separate planar and non-planar points which are further processed to generate segments via clustering. Buildings and trees were then separated by segment attributes. For instance, segments with small size (Vosselman et al., 2004), or segments with width shorter than 3 meters (Awrangjeb et al., 2010) were treated as trees. Segment-wise classification proved to be more reliable then point-wise methods. However, this technique usually requires high density of Lidar data which are not always available due to the high cost.

A number of research employed image information for separation of buildings and trees after initial segmentation using the Lidar data. The most frequently used information is NDVI (normalized difference vegetation index) estimated from multispectral images which are available in most of modern satellite and airborne sensors. A high NDVI value for a pixel indicates vegetation, whereas a low NDVI value generally indicates a non-vegetation pixel. While effective in most cases, NDVI has been found to be unreliable, particularly when non-vegetation pixels shared similar spectral attributes with vegetation. For instance, in the case when roofs have the similar color as trees, or trees have colors other than green (Awrangjeb et al., 2010). More complex methods exploited image textures. Image classification approaches using grey level co-occurrence matrix and self-organizing map classification have been investigated (Chen et al., 2006). These methods require large amount training samples, and are computationally expensive.

Existing approaches show varying degrees of success. Height data is effective in detection of low vegetation. Segmentation-based approaches with high density height data provide promising results, while low density point clouds are insufficient to reliably separate buildings and trees. Current image analysis methods mainly rely on pixel intensity, resulting omission and commission errors.

In this paper, we proposed a new image analysis approach based on texture and edge orientation derived from high resolution aerial imagery for enhanced classification of buildings and trees. The reported work is build upon the previous research, particularly the recent efforts described in Awrangjeb et al. (2010). In addition to high NDVI values, trees exhibit richer texture than building roofs. While buildings roofs may be painted in different colors, they usually have regular shape. The sides of the roofs are parallel to or perpendicular to each other. These texture and geometric property of buildings will be exploited to differentiate buildings and trees. The approach is detailed in the next Section.

3. APPORACH TO SEPARATION OF BUILDINGS AND TREES

A height threshold T_h=H_g+2.5, where H_g represents the ground height (DEM value), was applied to the raw Lidar data. This threshold removed objects of low height (shubbery, road furniture, cars, etc.) and preserved non-ground objects (trees and buildings). The outlines of the non-ground objects were extracted and the rectangle shapes were generated surrounding these objects using the techniques in Awrangjeb et al. (2010). NDVI was computed using multispectral imagery. Tree candidates were selected if their NDVI values were above the mean value of the NDVI. The rest of the non-ground objects were treated as buildings candidates. However, the use of NDVI was not always reliable, particularly in differentiation between trees and green buildings. In addition, some types of trees demonstrated low NDVI values. In fall or winter, many trees may become leafless, or the color of leaves change. These trees cannot be detected using NDVI, and will be misclassified as buildings. These two types of errors will be avoided using image texture information and edge orientation information, respectively, detailed as below.

3.1 Texture Analysis

We explore image entropy to identify green buildings from trees. Entropy is a statistical measure of randomness that can be used to characterize the texture of images (Gonzalez et al., 2003). Its adoption is based on the assumption that trees are rich in texture as compared to the roofs of buildings. A high entropy value indicates a texture (tree) pixel.

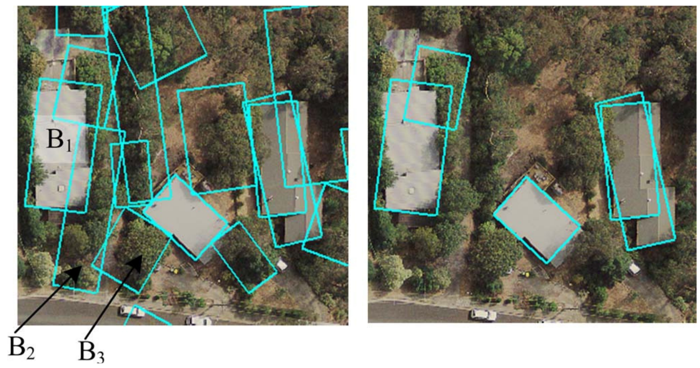
Entropy was calculated within a 9 by 9 window around a pixel. A normalized histogram H for the image window, involving 256 bins and values in the range of 0 to 1, was formed and entropy was calculated using non-zero frequencies as

$$e = -\sum H_i \log_2(H_i), \text{ where } 1 \leq i \leq 256 \text{ and } 0 \leq H_i \leq 1.$$

With the detected trees using NDVI, a further test is performed to check whether the average entropy is less than a predefined threshold. As the entropy values are generally low in buildings, green roofs which show similar colors as trees can be effectively identified.

3.2 Edge Orientation Analysis

As stated in the previous section, some trees do not appear green in aerial images, and are misclassified as buildings using NDVI. An example is given in Fig. 1where trees might not be pure green in color. Consequently, the method in Awrangjeb et al. (2010) produced a large number of false detection in the building candidates as shown in Fig. 1(a).



(a) (b)

Fig. 1 A complex scene with dense trees in a hilly terrain. (a) Detected building candidates using NDVI with a large number of false detection. (b) Detected buildings after removing false positives using edge orientation information.

Such errors can be avoided by exploiting object geometric property. Tree tops do not pose regular geometry as buildings. For instance, building edges are usually long and straight. They are parallel to or perpendicular to each other. On the other hand, tree edges are short, and their orientations are not arranged, demonstrating random distribution. We explored the edge orientation information to identify the trees which are misclassified as buildings using NDVI. This method was also used to confirm and validate the detected buildings.

With the detected candidate buildings using NDVI, a gradient histogram was formed using the edge points within each candidate building rectangle. Edges were first extracted from the image using an edge detector. Each edge $\Gamma(t)=(x(t),y(t))$ of length n, where t is an arbitrary parameter and $1 \leq t \leq n$, was smoothed by a Gaussian function g_σ with scale sigma σ:

$$x_\sigma(t) = x(t)*g_\sigma \text{ and } y_\sigma(t) = y(t)*g_\sigma$$

where * denotes convolution. Then, the first order derivatives $x'_\sigma(t)$ and $y'_\sigma(t)$ were calculated on the smoothed curve $\Gamma(t)=(x_\sigma(t),y_\sigma(t))$, and the gradient orientation can be estimated as

$$\Delta_\Gamma(t)= \arctan(y'_\sigma(t)/ x'_\sigma(t))$$

$\Delta_\Gamma(t)$ at each point will lie within the range of [-90º, +90º]. A histogram with a successive bin distance of 5º was then formed using the gradient orientation values of all edge points lying inside the candidate rectangle.

For buildings, one or more significant peaks should be observed in the gradient orientation histogram, since edges detected on building roofs were formed from straight line segments. All points on an apparent straight line segment will have a similar gradient orientation 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. Moreover, peaks separated by 90º correspond to perpendicular roof edges on buildings.

Fig. 2 illustrates three gradient orientation histogram functions and mean bin heights for candidate buildings B_1, B_2 and B_3 in Fig. 1(a). Fig. 2(a) shows that B_1 has two significant peaks: 80 pixels at 0º and 117 (55+62) pixels at 90º, these being well above the mean height of 28.6 pixels. The two significant peaks separated by 90º strongly suggest that this is a building. From Fig. 2(b) it can be seen that B_2 has one significant peak at 90º 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. 2(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 image resolution in this case was 10cm, so a bin height of 80 pixels indicates a total length of 8m from the contributing edges.

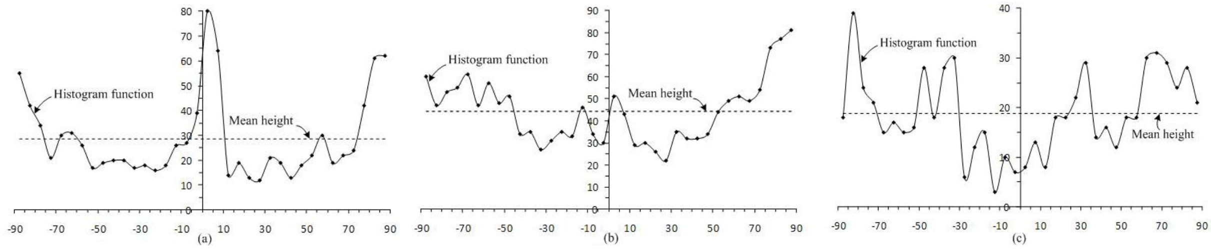


Fig. 2. Gradient orientation histogram functions and mean bin height for rectangles (a) B_1, (b) B_2 and (c) B_3 in Fig. 1 (a). The unit of horizontal axis is degree and the unit of vertical axis is pixel.

The observations above support the theoretical inferences. In practice, however, detected vegetation clusters may show the edge characteristics of a building, and a small building occluded by trees may not have sufficient edges to show the required peak properties. To overcome these problems, a set of rules was applied. If a detected rectangle passes at least one of the following tests it is selected as a building, otherwise it is treated as a tree.

Test 1: H has at least two peaks with heights of at least $3L_{min}$ (L_{min} is the minimum building length or width, set to 3m in our work) and the average height of bins between those peaks is less than L_{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.

Test 2: The highest bin in His at least $3L_{min}$ in height and the aggregated height of all bins in H 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 trees where the aggregated height of all bins is high.

Test 3: H has at least two peaks with heights of at least $2L_{min}$, and the highest bin to mean height ratio 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.

Test 4: The highest bin in H has a height of at least L_{min} and the highest bin to mean height ratio 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.

The application of these tests on the complex scene in Fig. 1(a) produced the results in Fig. 1(b). Note that this rule-based procedure using edge orientation effectively removed the false candidates, and buildings were correctly detected.

4. EXPERIMENTS

The developed approaches have been tested with different datasets over varying terrain types. The test sites include three suburban areas in Australia, Fairfield in New South Wales, Moonee Ponds and Knox in Victoria. There are 370 buildings, 250 buildings, and 130 buildings in Fairfield, Moonee Ponds and Knox datasets, respectively. Fairfield contains many large industrial buildings and in Moonee Ponds the roofs of some buildings appear green in the images. Knox can be characterized as an outer suburban with lower housing density and extensive tree coverage that partially occluded buildings. In terms of topography, Fairfield and Moonee Ponds are relatively flat while Knox is quite hilly. Lidar coverage comprised last-pulse returns with a point spacing of 0.5m for Fairfield, and first-pulse returns with a point spacing of 1m for Moonee Ponds and Knox. For Fairfield and Knox, RGB color orthoimagery was available, with resolutions of 0.15m and 0.1m, respectively. Moonee Ponds image data comprised RGBI color orthoimagery with a resolution of 0.1m. Bare-earth DEMs of 1m horizontal resolution covered all three areas, and were used to generate orthoimagery. Therefore, the building roofs and the tree-tops were displaced with respect to the Lidar data, and thus, data alignment was not perfect.

The results were evaluated using manually collected reference data which were created by monoscopic image measurements. All rectangular structures, recognizable as buildings were digitized. The reference data included garden sheds, garages, etc. These were sometimes as small as $10m^2$ in the areas. For performance assessment, completeness and correctness measures (Awrangjeb et al., 2010) are employed.

Table 1 shows performance evaluation of the results obtained for the three datasets with our approach. Visual illustrations of the detection results are shown in Fig 3. Compared with the results derived from the algorithms proposed in Awrangjeb et al. (2010), our approach produced moderately better performance within both Fairfield and Moonee Ponds. The better performance was mainly due to proper detection of large industrial buildings in Fairfield, detection of some green buildings using image texture in Moonee Ponds, and elimination of trees with edge orientation in both Fairfield and Moonee Ponds.

Table 1. Performance evaluation.

	Fairfield	Moonee Ponds	Knox	**Average**
Completeness (%)	95.1	94.5	93.2	94.0
Correctness (%)	95.4	95.3	87.2	91.3

(a)

(b)

(c)

Fig. 3. Separation of trees and buildings for building detection in (a) Fairfield, (b) Moonee Ponds and (c) Knox. In (c), the detected buildings by Awrangjeb et al. (2010) on two samples are shown in left for comparison, while the detected buildings with the methods described here are presented on the right.

In Knox, our approach also performed very well even if the scene is very complex with hilly terrain and dense tall trees which occlude significantly buildings. For comparison, the scene images were also processed with the methods described in Awrangjeb et al. (2010) with the results shown on the right of Fig. 3(c). It can be observed that significant improvement has been achieved. Awrangjeb's method generated a large number of false detections in Knox, and some buildings were not detected, as illustrated in the left of Fig. 3(c). Consequently, only 77% completeness and 67%correctness were observed. This is because the method is not very effective in differentiating buildings and trees, particularly when the imagery lacks near infrared information and the NDVI was computed using red and green bands. In contrast, as shown for Knox on the right of Fig. 3(c), our approach picked up the buildings and removed a large number of false positives using its gradient orientation histogram, significantly improving the results. The completeness and correctness increased to over 93% and 87%, respectively. In general, our approach offered on average across the three datasets a more than 10% increase in completeness and correctness.

5. CONCLUSIONS

This paper presented a new approach to efficient separation of buildings and trees for improved building detection. Lidar data were firstly employed to remove low vegetations and detect above ground objects including trees and buildings. Trees and buildings were then initially differentiated with NDVI. New approaches were proposed to avoid omission and commission errors. Firstly, texture analysis with image entropy identified buildings with similar color as trees. Trees, which were misclassified as buildings, were detected with rule-based approach using edge orientation histogram information. These methods significantly improved the success rate of the building detection as demonstrated in the test data with varying terrains and land covers. Compared with other methods, the proposed approaches achieved more than 10% increase in completeness and correctness. In particular, our method proved to be very effective in densely vegetated areas which is a challenge in most building detection methods.

It is acknowledged that there will be situations in which the developed algorithm will fail. For example, textured green roofs may not be distinguished from trees using the entropy information. In addition, trees with shadows and self-occlusions display very low entropy values, and thus may be misclassified as buildings using entropy information. While such error might be avoided with edge orientation histogram, the parameter must be carefully set in the rule-based procedure. Our current research focuses upon resolving these problems as well as upon the 3D reconstruction of complex building roofs.

ACKNOWLEDGEMENTS

The authors would like to thank the Department of Sustainability and Environment, Victoria, Australia for providing the Lidar data and orthoimagery for the Moonee Ponds and Knox test sites.

REFERENCES

Awrangjeb, M., Ravanbakhsh, M., Fraser, C.S., 2010. Automatic detection of residential buildings using lidar data and multispectral imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 65(5): 457-467.
Chen, L., Teo, T., Hsieh, C., Rau, J., 2006. Reconstruction of building models with curvilinear boundaries from laser scanner and aerial imagery. Lecture Notes in Computer Science, 4319, 24-33.
Dash, J., Steinle, E., Singh, R.P., Bähr, H.P., 2004. Automatic building extraction from laser scanning data: an input tool for disaster management. Advances in Space Research, 33(3):317-322.
Khoshelham, K., Nedkov, S., Nardinocchi, C., 2008. A comparison of bayesian and evidence-based fusion methods for automated building detection in aerial data. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 37(part B7), pp. 1183-1188.
Maas, H.~G., 2001. The suitability of airborne laser scanner data for automatic 3d object reconstruction. In: Proc. 3rd International Workshop on Automatic Extraction of Man-Made Objects from Aerial and Space Images. Ascona, Switzerland, pp. 345--356.
Mayer, H., 1999. Automatic object extraction from aerial imagery - a survey focusing on buildings. Computer Vision and Image Understanding, 74(2):138-149.
Rottensteiner, F., Trinder, J., Clode, S., 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):135-149.
Sampath, A. and Shan, J., 2010. Segmentation and Reconstruction of Polyhedral Building Roofs From Aerial Lidar Point Clouds. IEEE Transactions on Geoscience and Remote Sensing, Vol. 48(3):1554-1567.
Sohn, G., 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): 43-63.
Vosselman, G., Gorte, B.G.H, Sithole, G., Rabani, T., 2004. Recognizing structure in laser scanner point clouds. International Archives of Photogrammetry, Remote Sensing and Spatial Information Science, Vol. 46, Part 8/W2, Freiburg, German, 4-6 October, pp. 33-38.
Vu, T., Yamazaki, F., Matsuoka, M., 2009. Multi-scale solution for building extraction from lidar and image data. International Journal of Applied Earth Observation and Geoinformation, 11(4): 281-289.
Zhang, K., Yan, J., Chen, S.C., 2006. Automatic construction of building footprints from airborne lidar data. IEEE Transactions on Geoscience and Remote Sensing, 44(9): 2523-2533.