Detecting Buildings and Roof Segments by Combining LIDAR Data and Multispectral Images

Franz Rottensteiner[1], John Trinder[1], Simon Clode[2], and Kurt Kubik[2]
[1]School of Surveying and Spatial Information Systems,
The University of New South Wales, Sydney, NSW 2052, Australia
{f.rottensteiner, j.trinder}@unsw.edu.au
[2]The Intelligent Real-Time Imaging and Sensing Group, School of ITEE
The University of Queensland, Brisbane, QLD 4072, Australia
{sclode, kubik}@itee.uq.edu.au

Abstract

A method for the automatic detection of buildings and their roof planes from LIDAR data and multispectral images is presented. For building detection, a classification technique is applied in a hierarchic way to overcome the problems encountered in areas of heterogeneous appearance of buildings. The detection of roof planes is based on a region growing algorithm applied to the LIDAR data, the seed regions detected by a grey-level segmentation of the multispectral images. We describe the algorithms involved, giving examples for a test site in Fairfield (Sydney).

Keywords: airborne laser scanning, building detection, data fusion

1 Introduction

1.1 Motivation and Goals

In addition to photogrammetric techniques relying on aerial images, the generation of 3D building models from point clouds provided by airborne laser scanning, also known as LIDAR (*LI*ght *D*etection *A*nd *R*anging), is gaining importance. This development has been triggered by the progress in sensor technology which has rendered possible the acquisition of very dense point clouds using airborne laser scanners [5]. Using high-resolution LIDAR data, it is not only possible to detect buildings and their approximate outlines, but also to extract planar roof faces and, thus, to create models correctly resembling the roof structure [1], [7], [10]. In order to do so, buildings first have to be detected in the input data. After that, the buildings are reconstructed by polyhedral models, which requires the roof planes to be found in the regions of interest.

With decreasing resolution, it becomes more difficult to correctly reconstruct buildings from LIDAR data. In order to improve the performance of building reconstruction, additional data can be considered. First, LIDAR systems register two echos of the laser beam, the *first* and the *last pulse*, corresponding to the highest and the lowest object point hit by the laser beam. First and last pulse data will especially differ in the presence of vegetation [5]. Second, along with the run-time of the signal, the *intensity* of the returned laser beam is registered by LIDAR systems. LIDAR systems typically operate in the infrared part of the electromagnetic spectrum. Unfortunately, the intensity image is undersampled and, thus, very noisy [11]. Third, *multispectral images* can provide valuable information due to their spectral content and because their resolution is still better than the resolution of laser scanner data [9].

Building reconstruction is also made complicated by the fact that buildings may have quite different appearances both with respect to their geometric dimensions and their reflectance properties. Thus, it is often impossible to select appropriate thresholds and filter sizes for various algorithms. It is a well-known strategy in image matching to apply algorithms to data having a lower resolution first to get approximate values, refining these results in an iterative way, in each iteration considering data of a better resolution than in the previous one. In this paper, we first want to describe a method for building detection from LIDAR data and multispectral images making use of such a hierarchic approach. Second, we want to present a new iterative method for combining colour images and roof planes in the building regions thus detected. The examples presented in this paper were computed using the LIDAR and image data from a test site in Fairfield (Sydney).

1.2 Related Work

There have been several attempts to detect buildings in LIDAR data in the past. The task has been solved by classifying the LIDAR points according to whether they belong to the terrain, to buildings or to other object classes. Morphological opening filters or rank filters are used to determine a digital terrain model

(DTM) which is subtracted from the digital surface model (DSM). By applying height thresholds to the normalized DSM thus created, an initial building mask is obtained [12]. The initial classification has to be improved in order to remove vegetation areas. In [2], this is accomplished by a framework for combining various cues in a Bayesian network. The problems with their approach are related to the complexity of estimating the conditional probabilities required for the Bayesian network.

In [6], a DSM derived by image matching and a colour image are fused on the basis of the Dempster-Shafer theory. Fusion is carried out at feature level, the initial segmentation being performed by a K-means unsupervised classification of the colour images, using shape cues such as the normalised difference vegetation index (NDVI) and the average relative height of the feature to distinguish buildings from other objects.

In [7], we have presented an algorithm for building detection that relied on DTM generation by hierarchic robust linear prediction [3], using the DTM and DSM grids for further classification. The method gives good results in densely built-up areas [7], but in more heterogeneous areas containing houses of different sizes and also forests, tuning the parameters for DTM generation is not an easy task. In this paper, we will describe how that method has been modified to consider the additional data sources and to work in a hierarchic way without relying on hierarchic robust linear prediction.

Roof planes can be detected by segmenting the LIDAR data. In [10], a method based on the Hough-transformation is used for that purpose. Schenk and Csatho [9] proposed feature based fusion of LIDAR data and digital aerial images for obtaining a better surface description than it can be achieved by using only one of these data sources. We have presented an iterative method for LIDAR segmentation in [7] which delivered the most salient roof planes from LIDAR data in a densely built-up urban area. That method is based on region growing, the seed regions being detected by an analysis of the second derivatives of a LIDAR DSM. The method also requires a certain minimum number of LIDAR points per roof plane, a condition that is often not fulfilled for smaller buildings. In this paper, we want to present a similar method that uses the results of a segmentation of a digital image of higher resolution than the LIDAR data to detect seed regions for region growing, whereas region growing is performed by analysing the planar fit of the points in the LIDAR DSM.

2 Building Detection

The work flow for our method for building extraction is depicted in figure 1. The input to our method is given by three data sets that have to be generated from the raw data in a preprocessing step. The *last pulse DSM* is sampled into a regular grid by linear prediction using a straight line as the covariance function, so that the interpolation is carried out almost without filtering [7]. The *first pulse DSM* is also sampled into a regular grid. The *normalised difference vegetation index (NDVI)* is computed from the near infrared and a red band of the multispectral images we assume to be available [6]. These image data must be geocoded.

First, a DTM is created from the last pulse DSM by morphological grey scale opening using a square structural element. The DTM does not contain objects smaller than the structural element. If the structural element is chosen to be greater than the largest building in the data set, all buildings will be removed; however, if it is too large, terrain details might be removed, too. If the structural element is chosen small, the results of morphological opening will be closer to the DSM and, thus, to the terrain in areas where the DSM represents the terrain, but larger buildings will remain in the data set. This is the reason why we apply morphological filtering in a hierarchical framework. Initially, we select the size of the structural element to be the size of the largest building available in the dataset.

An initial building mask is created mainly by thresholding operations. Morphological opening is used to eliminate spurious and strangely shaped areas, and connected components of “building” pixels are considered to be the initial building regions. For these regions, we evaluate the surface roughness and the average NDVI. In the early iterations, very tight thresholds are applied to surface roughness, because we assume the largest buildings in the scene to consist of large planar surfaces. Thus, we obtain an intermediate set of building regions, only containing the largest buildings (corresponding to the current state of the DTM). After that, the procedure is repeated with a smaller structural element for morphological opening, but in the areas already classified as buildings, the DTM heights of the previous iteration are substituted for the results of the morphological filter. Thus, the buildings already detected are eliminated, whereas the smaller size of structural element for morphological filtering helps to obtain a finer approximation for the DTM, rendering possible the separation of smaller buildings. The whole procedure is repeated with a decreasing size of the structural element until a user-specified minimum size is reached. The results of the final iteration are identical to the results of building detection, basically represented by a list of building regions and a building label image.

Detecting Candidate Regions: Morphological filtering provides us with an approximation for the DTM. For buildings aready detected, the DTM generated by morphological opening in the previous iteration is substituted for the results of morphological filtering, so that large buildings that would be preserved

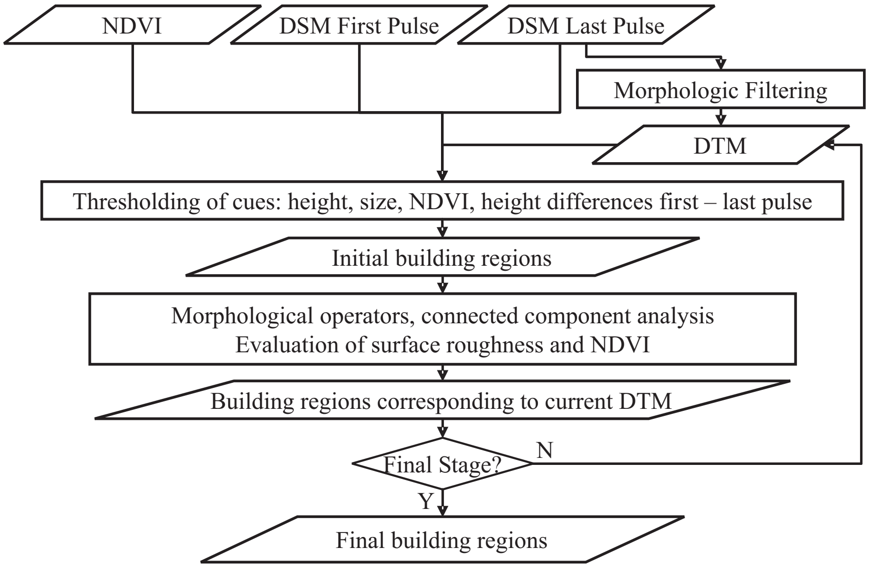

Figure 1: The work flow for building detection.

by morphological filtering in the current iteration are eliminated beforehand. An initial building mask is created by thresholding the height differences between the last pulse DSM and the DTM (e.g., by h_{min}=2.5 m). This initial building mask still contains spurious regions, areas covered by vegetation, and terrain structures smaller than the structural element for morphological filtering.

At this instance, the additional data sources can be used to improve these results. First, a large NDVI indicates areas covered by vegetation, so that pixels having an NDVI above a certain threshold are erased in the building mask. Second, the heights from first and last pulse data will differ mainly in areas covered by trees and if the laser beam accidentally hits the roof edge of a building. Thus, in most cases, large height differences between first and last pulse data indicate trees. Pixels having a height difference larger than a certain threshold are erased in the building mask, too.

A binary morphological opening filter using a structural element of a size corresponding to the expected minimum size of a building part (e.g., $3 \times 3\ \mathrm{m}^2$) is applied to the initial building mask to erase oddly shaped objects such as fences and to separate building regions just bridged by a thin line of pixels. A connected component analysis of the resulting image is applied to obtain the initial building regions. Regions smaller than a minimum area are discarded.

Classification of Candidate Regions: Some of the initial building regions correspond to groups of trees or to terrain structures smaller than the structural element. These regions can be eliminated by evaluating a surface roughness criterion derived by an analysis of the second derivatives of the DSM. In [4], a method for polymorphic feature extraction is described which aims at a classification of texture as being homogeneous, linear, or point-like, by an analysis of the first derivatives of a digital image. This method is applied to the first derivatives of the DSM using an integration kernel of a size corresponding to, e.g., 5 m in object space. Under these circumstances, “homogeneous” pixels correspond to areas of locally parallel surface normal vectors, thus, they are situated in a locally planar neighbourhood. “Linear” pixels correspond to the intersections of planes, and “point-like” pixels are in a neighbourhood of great, but anisotropic variations of the surface normal vectors. For evaluating surface roughness, the numbers of “homogeneous” and “point-like” pixels are counted in each initial building region. Buildings are characterised by a large percentage of “homogeneous” and by a small percentage of “point-like” pixels. By comparing these percentages to thresholds, non-building regions can be eliminated. The surface roughness criterion works well for large buildings and with dense LIDAR data [7]. If the point distance of the LIDAR data is larger than, e.g., 1 m, only few LIDAR points will be situated on small buildings, so that the percentage of “homogeneous” pixels is reduced, whereas the percentage of “point-like” pixels is increased. Thus, it makes sense to additionally evaluate the average NDVI for each building region to eliminate vegetation areas.

Finally, vegetation areas still connected to buildings are eliminated. By morphological opening, regions just connected by small bridges are separated. The resulting binary image is analyzed by a connected component analysis which results in a greater number of regions, and the terrain roughness criterion is evaluated again. Regions now classified as containing vegetation are erased in the initial building label image [7].

3 Detection of Roof Segments

In [8], we described a method for the detection of planar roof segments in high-resolution LIDAR data. The method was also based on polymorphic feature extraction [4]. We used connected components of pixels classified as "homogeneous", thus being in a local neighbourhood of parallel surface normals, as seed regions for region growing. In order to overcome classification errors in polymorphic feature extraction, seed region selection was applied to the LIDAR data in an iterative way, using different significance levels for the classification. First, only the "most significantly planar" pixels of the DSM were used as seed regions, and region growing was only applied to seed regions having a good r.m.s. error of planar fit. After that, the classification was repeated, using another significance level for classification and applying our procedure for seed region selection to connected components of homogeneous pixels not yet assigned to a planar segment. In this way, a certain number of iterations was carried out, in each iteration taking into account less and less significantly homogeneous areas.

This procedure requires a certain minimum number of DSM pixels corresponding to a roof plane. For small buildings and LIDAR point distances larger than 1 m, the roof planes cannot be separated by the method: In the early iterations, too small a number of pixels will be classified as "homogeneous" because in any local neighbourhood, there will be a certain variation of the surface normals, whereas in later iterations, intersection lines of roof planes will be missed and, thus, roof planes will be merged. It is the idea of the new algorithm presented in this paper to select the seed regions in digital images having a better resolution than the LIDAR data. These seed regions are propagated to the LIDAR DSM, where the parameters of the corresponding planes can be computed, and region growing is applied in the way described above. Again, the procedure of seed region selection and region growing has to be carried out in an iterative way to overcome problems related to grey level segmentation of the digital images.

Detection of Seed Regions in the Digital Image: Seed region selection is again based on polymorphic feature extraction, but this time applied to a digital image. We only use the brightness of the images for that purpose. We assume connected components of pixels classified as "homogeneous", thus, being in a local neighbourhood of homogeneous brightness, as seed regions. Unfortunately, due to the lack of contrast of neighbouring roof planes, most of these seed regions will contain more than one relevant roof plane. We reduce the size of the seed regions by shrinking operations, and we further split them into smaller parts by morphologic filtering. Thus, we get a larger number of seed regions that are also smaller than the original ones, so that we can assume that at least some of them will correspond to a single roof plane in object space.

Region Growing in the LIDAR DSM: The seed regions are propagaged to the LIDAR DSM. In case the digital image is an orthophoto, this just means that the DSM heights have to be interpolated at the positions of the orthophoto pixels. From these interpolated DSM points, the plane parameters of the seed regions can be estimated, and region growing can be applied to the planes having a good planar fit indicated by a small r.m.s. error, adding a DSM grid point at the border of a planar segment to that segment if the point's distance from the adjusting plane is below a certain threshold. Region growing is performed iteratively, each time checking all DSM pixels neighbouring the segment and not yet belonging to another segment. After each iteration, the plane parameters are re-computed. Iteration stops if no more pixels can be added to any planar segment.

Hierarchic Roof Segment Detection: Roof plane detection is carried out in an iterative way. First, it is iterated over different significance levels of classification in polymorphic feature extraction, thus computing the seed regions from less and less significantly homogeneous pixels. At each stage of polymorphic feature extraction, there is a further loop of iterations, this time by using different filter sizes for reducing the sizes of the initial seed regions: Initially, the seed regions have to be reduced considerably in size, so that many planes will be eliminated completely; otherwise, roof planes would be merged. However, if region growing succeeds in extracting at least some of the roof planes, seed region detection can be repeated, just considering the pixels not yet belonging to a roof plane already detected. The seed regions will still be reduced in size, but in each iteration, this size reduction will be a smaller one, because in each iteration, less and less pixels will remain unclassified. In this way, it is possible to obtain a segmentation better than any that could be obtained by using just one of the data sources.

4 Experiments

The test data set was captured over Faifield (Sydney) using an Optech ALTM 3025 laser scanner. Both first and last pulse were recorded, the average point distance being about 1.2 m. We derived DSM grids at a resolution of 1 m from these data (left part of figure 2). Intensity data were available, too. We used an area of 2×2 km^2 for our test. For that area, a true colour digital orthophoto with a resolution of 15 cm was available. The orthophoto was created using a DTM, so that the roofs and the tree-tops were slightly displaced with respect to the LIDAR data. Unfortunately, the digital orthophoto did not contain an infrared band. We circumvented this problem by resampling both the digital orthophoto and the (infrared) LIDAR intensity data to a resolution of 0.5 m and by computing a “pseudo-NDVI-image” from the LIDAR intensities and the green band of the digital orthophoto (right part of figure 2). Due to problems with georeferencing and shadow areas, the “pseudo-NDVI-image” did not provide as much information as the NDVI image would have.

Figure 2: The test site. Left: DSM (black: low areas, white: high areas). Total area: $2000 \times 2000\,\mathrm{m}^2$. Right: The “pseudo-NDVI-image”.

4.1 Building Detection

We started the procedure by morphological opening of the DSM using a structural element of $150 \times 150\ \mathrm{m}^2$. Using that structural element, all buildings can be eliminated, but the terrain shape is not well preserved, so that the residential buildings in the lower part of the scene are merged. Using a smaller structural element, more terrain details are preserved, but the large buildings are still contained in the data. However, using our hierarchic approach, these buildings can be eliminated beforehand, as described in section 2.

The initial building mask was derived using a height threshold of 2 m. Texture classification was carried out using a filter kernel size of 3×3 pixels. In the first iteration, starting from the DTM created by using a structural element of $150 \times 150\ \mathrm{m}^2$, we only accepted regions larger than $2500\ \mathrm{m}^2$, containing less than 0.30% of “point-like” and at least 70% of “homogeneous” pixels, thus, large regions consisting of few, but large roof planes. As the industrial buildings had a high reflectivity in the infrared part of the spectrum, the threshold for the “pseudo-NDVI” was kept at 75%. 85 mostly large building structures are detected in the first iteration.

Altogether three iterations were carried out, using structural elements of 150 m, 75 m, and 25 m. In the final iteration, regions larger than $25\ \mathrm{m}^2$, containing less than 85% of “point-like” and at least 1% of “homogeneous” pixels were accepted. These relatively loose settings of the threshold were a consequence of the LIDAR resolution, with only few points and, thus, few “homogeneous” pixels on the roofs of the smaller buildings. 1589 buildings were detected (figure 3). The parameters for classification were chosen in a way to minimise the number of missed buildings, accepting a larger rate of false alarms. Less than 1% of the buildings were not detected. There remain some trees in the data, and some of the buildings are still merged, especially if the distance between them is small. As almost all the buildings are contained in the data, it might be possible to improve the results of classification by considering additional cues derived, for instance, from the colour images.

Figure 3: Buildings extracted in the third iteration.

4.2 Detection of Roof Segments

Figure 4 shows the brightness of the digital orthophoto and the DSM for one of the buildings detected in the Fairfield data set. The orthophoto was used at its full resolution (15 cm), and the DSM was interpolated at 0.5 m. The homogeneous segments derived by polymorphic feature extraction using two different significance levels for classification of the brightness image are displayed in figure 5. It is obvious that some relevant planes cannot be correctly separated due to the lack of contrast between them. By shrinking the homogeneous segments in the left part of figure 5, small seed regions can be generated, and applying region growing to the seed regions having a r.m.s. error of planar fit better than 15 cm results in the 5 planes detected in the left part of figure 6. In the first iteration, the seed regions are shrinked by 11 pixels and then

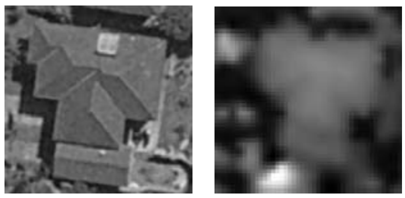

Figure 4: One of the buildings in the test dataset. Left: Digital orthophoto (resolution: 15 cm). Right: DSM.

morphologically opened using a structural element of the same size. This size is reduced by 2 in each iteration. Altogether 5 iterations are carried out in this way for each stage of polymorphic feature extraction, and four such stages are considered. The final results are presented in the right part of figure 6. Despite of the resolution of the LIDAR data and despite of the fact that the roof planes could not be separated in the image segmentation, the relevant roof planes could be detected. There are still some spurious segments, and the segment outlines are not yet correctly shaped. The latter can be improved by computing the intersection lines of neighbouring segments [8].

Figure 5: Homogeneous segments derived for two different significance levels.

Figure 6: Roof segments. Left: first iteration. Right: last iteration.

5 Conclusion and Future Work

We have presented a hierarchic method for building detection from LIDAR data and multispectral images, and we have shown its applicability in a test site of heterogeneous building shapes. In our test, we put more emphasis on detecting all buildings in the test data set than on reducing the false alarm rate because in the future we want this method to be the module for initial segmentation in a framework using more sophisticated methods of data fusion similar to those described in [6]. Using the NDVI computed from real infrared images rather than the "pseudo-NDVI" used in this test might help to improve the results. We have also presented a new method for combining digital images and LIDAR data for the detection of roof planes. The method is still work in progress. We especially want to reduce the number of parameters involved in the process. Our preliminary results show that, using this method, it was possible to detect more roof planes than from one of the data sources alone.

6 Acknowledgements

This work was supported by the Australian Research Council (ARC) under Discovery Project DP0344678. The Fairfield data set was provided by AAM Geoscan, Kangaroo Point, QLD 4169, Australia (www.aamgeoscan.com.au).

References

[1] Brenner, C.: Dreidimensionale Gebäuderekonstruktion aus digitalen Oberflächenmodellen und Grundrissen. PhD Thesis, Inst. of Photogrammetry, Stuttgart University, DGK-C 530 (2000).

[2] Brunn, A., Weidner, U.: Extracting buildings from digital surface models. IAPRS **XXXII**(3-4W2) (1997) 27–34.

[3] Briese, C., Pfeifer, N., Dorninger, P.: Applications of the robust interpolation for DTM determination. IAPSIS **XXXIV**/3A (2002) 55–61.

[4] Fuchs, C.: Extraktion polymorpher Bildstrukturen und ihre topologische und geometrische Gruppierung. PhD Thesis, Inst. of Photogrammetry, Bonn University, DGK-C 502 (1998).

[5] Kraus, K.: Principles of airborne laser scanning. Journal of the Swedish Society for Photogrammetry and Remote Sensing **2002**:1 (2002) 53–56.

[6] Lu, Y. H. and Trinder, J.: Dempster-Shafer data fusion theory for automatic extraction buildings in 3D reconstruction. ASPRS Workshop Anchorage, Alaska (2003).

[7] Rottensteiner, F. and Briese, C.: A new method for building extraction in urban areas from high-resolution LIDAR data. IAPSIS **XXXIV**/3A (2002) 295–301.

[8] Rottensteiner, F., and Briese, C.: Automatic generation of building models from LIDAR data and the integration of aerial images. Accepted for the ISPRS workshop on "3-D reconstruction from airborne laserscanner and InSAR data" in Dresden, Germany, Oct. 2003. To appear in IAPRSIS.

[9] Schenk, T. and Csatho, B.: Fusion of LIDAR data and aerial imagery for a more complete surface description. IAPSIS **XXXIV**/3A (2002) 310–317

[10] Vosselman, G., Dijkman, S.: 3D building model reconstruction from point clouds and ground plans. IAPSIS **XXXIV**(3W4) (2001) 37–43.

[11] Vosselman, G.: On the estimation of planimetric offsets in laser altimetry data. IAPSIS **XXXIV**/3A (2002) 375–380.

[12] Weidner, U.: Gebäudeerfassung aus digitalen Oberflächenmodellen. PhD Thesis, Inst. of Photogrammetry, Bonn University, DGK-C 474 (1997).