FUSING AIRBORNE LASER SCANNER DATA AND AERIAL IMAGERY FOR THE AUTOMATIC EXTRACTION OF BUILDINGS IN DENSELY BUILT-UP AREAS

F. Rottensteiner [a, *], J. Trinder [a], S. Clode [b], K. Kubik [b]

[a] School of Surveying and Spatial Information Systems, The University of New South Wales, Sydney, NSW 2052, Australia - {f.rottensteiner, j.trinder}@unsw.edu.au
[b] The Intelligent Real-Time Imaging and Sensing Group, School of ITEE, University of Queensland, Brisbane, QLD 4072, Australia - {sclode, kubik}@itee.uq.edu.au

Commission III, WG III/6

KEY WORDS: Building, Extraction, LIDAR, Multisensor, Fusion

ABSTRACT:

Using airborne laser scanner data, buildings can be detected automatically, and their roof shapes can be reconstructed. The success rate of building detection and the level of detail of the resulting building models depend on the resolution of the laser scanner data, which is still lower than the resolution of aerial imagery. Building extraction from aerial images alone is difficult because of problems related to shadows, occlusions, and poor contrast. That is why it makes sense to combine these data sources to improve the results of building extraction. This article deals with the fusion of airborne laser scanner data and multi-spectral aerial images for building extraction. There are three instances in the overall process when exploiting the complementary properties of these data appears to be most beneficial: building detection, roof plane detection, and the determination of roof boundaries. Building detection is based on the Dempster-Shafer theory for data fusion. In roof plane detection, the results of a segmentation of the laser scanner data are improved using the digital images. The geometric quality of the roof plane boundaries can be improved at step edges by matching the object edges of the polyhedral models with image edges. Examples are given for a test site in Fairfield (NSW).

1. INTRODUCTION

1.1 Motivation and Goals

The high potential of LIDAR (*LI*ght *D*etection *A*nd *R*anging) data for automatic building extraction has been shown in the past, e.g. (Vosselman and Dijkman, 2001). The success rate of building detection and the level of detail of the resulting building models depend on the resolution of the laser scanner data, which is typically still lower than the resolution of aerial imagery. On the other hand, building extraction from aerial images alone is difficult because of shadows and occlusions, and also because the transition from 2D image space to 3D object space has to be carried out. That is why it makes eminent sense to combine these data sources to improve the results of building extraction. There are three instances when exploiting the complementary properties of these data appears to be most beneficial (Rottensteiner and Briese, 2003):

(1) Building detection: The main problem in this context is to distinguish buildings from trees. LIDAR data give parameters of surface roughness, but with decreasing resolution of the LIDAR data, the classification becomes more critical in areas where the appearance of trees and buildings is similar. The height differences between the first and the last echoes of the laser pulse and multi-spectral images can be used as additional data sources to improve the classification results.
(2) Roof plane detection: In order to reconstruct the buildings by polyhedral models, roof planes have to be detected first. Large roof planes can be detected in the LIDAR data. The results thus achieved can be improved by taking into account aerial images.

(3) Determination of roof boundaries: The geometric quality of the roof boundaries at step edges, which in general is poor if only LIDAR data are used, can be improved by image edges.

This paper deals with the fusion of first and last pulse LIDAR data and multi-spectral aerial images for building extraction. It consists of two main parts. The first part, describing our method for building detection and the results that could be achieved by it, is presented in section 2. In the second part, which is presented in section 3, we will describe the current state of our work for the fusion of LIDAR and image data for roof plane detection and the determination of roof boundaries. Section 4 gives conclusions and an outlook on future work.

1.2 Related work

1.2.1 Building detection: The building detection starts with the generation of a coarse digital terrain model (DTM) from the digital surface model (DSM) provided by the LIDAR data, e.g. by morphologic filtering. A further classification must separate points on buildings from points on trees and other objects by evaluating the local surface roughness and other cues. With multi-spectral images, the normalised difference vegetation index (NDVI) is well-suited for classification in this context (Lu and Trinder, 2003).

Various classification techniques have been applied for building detection, e.g., unsupervised classification (Haala and Brenner, 1999), rule-based classification (Rottensteiner et al., 2003), Bayesian networks (Brunn and Weidner, 1997), and fuzzy logic (Voegtle and Steinle, 2003). The probabilistic approaches among the cited ones face the difficulty of modelling the

* Corresponding author.

a priori probabilities without training samples, which is problematic if the assumption of a normal distribution of the data vectors is unrealistic. We propose to use the theory of Dempster-Shafer for data fusion, because its capability of handling incomplete information gives us a tool to reduce the degree to which we have to make assumptions about the distribution of our data (Klein, 1999; Lu and Trinder, 2003).

1.2.2 Roof plane detection and delineation: Ameri and Fritsch (2000) combined a DSM and aerial images for the geometrical reconstruction of buildings by polyhedrons. They searched for co-planar pixels in the DSM, which resulted in seed regions for region growing in one of the aerial images. The resulting roof planes were combined to form a polyhedral model, which was then improved by fitting the model to image edges. Problems were mainly caused by poor contrast, because region growing was only applied to one of the aerial images, and because the 3D information provided by the DSM was not included in the region growing process.

Schenk and Csatho (2002) put forward the idea of exploiting the complementary properties of LIDAR data and aerial images to achieve a more complete surface description by feature based fusion. LIDAR data are useful for the detection of surface patches having specific geometrical properties and for deriving parameters related to surface roughness, whereas aerial images can help to provide the surface boundaries and the locations of surface discontinuities. The planar patches detected in LIDAR data are used to improve the results of edge detection in the aerial images, and the image edges thus extracted help to improve the geometrical quality of the surface boundaries.

Rottensteiner and Briese (2003) described a method for roof plane detection from LIDAR data, and they discussed strategies for integrating aerial images in their work flow for building reconstruction. They proposed to improve their initial planar segmentation by adding new planar segments to the original ones if sufficient evidence is found in the aerial images. They also presented an adjustment model for wire-frame fitting. In (Rottensteiner et al., 2003), we have shown how planar segments can be detected by a combined segmentation of a digital orthophoto and a DSM. In this work, we want to show how the initial segmentation of the DSM can be improved by matching the planar patches with homogeneous regions extracted from two or more aerial images. This will result in better approximations for the roof boundaries, and it will support the distinction between roof plane intersections and step edges (i.e. intersections between roof planes and walls). We also want to show how the geometric quality of the step edges can be improved using edges extracted from the digital images.

1.3 The Test Data Set

Our test data were captured in Fairfield (NSW), covering an area of 2 x 2 km^2. The LIDAR data were captured using an Optech laser scanner. Both first and last pulses and intensities were recorded with an average point distance of about 1.2 m. We derived DSM grids at a resolution of 1 m from these data. True colour aerial stereo images (1:11000, f = 310 mm) were also available. These images were scanned at a resolution of 15 μm, corresponding to 0.17 m on the ground. A digital orthophoto with a resolution of 0.15 m was created using a DTM. Unfortunately, the digital images did not contain an infrared band, which would have been necessary for computing the NDVI. We circumvented this problem by resampling both the digital orthophoto and the LIDAR intensity data (wavelength: 1064 nm) to a resolution of 1 m and by computing a "pseudo-NDVI-image" from the LIDAR intensities and the red band of the digital orthophoto.

In order to evaluate the results of building detection, a reference data set was created by digitising building polygons in the digital orthophoto. We chose to digitize all structures recognisable as buildings independent of their size. The reference data include garden sheds, garages, etc, that are sometimes smaller than 10 m^2 in area. Neighbouring buildings that were joined, but are obviously separate entities, were digitized as separate polygons, and contradictions between image and LIDAR data were excluded. Thus, altogether 2337 polygons could be used for evaluation.

2. BUILDING DETECTION

The input to our method for building detection is given by three data sets. The *last pulse DSM* is sampled into a regular grid by linear prediction with a low degree of filtering. The first pulse DSM is also sampled into a regular grid, and by computing the height differences between these DSMs, we obtain a model of the *height differences between the first and the last pulses* ΔH_{FL}. The *normalised difference vegetation index (NDVI)* is computed from the near infrared and the red bands of a geocoded multi-spectral image (Lu and Trinder, 2003).

The work flow for our method for building detection consists of two stages. First, a coarse DTM has to be generated. We use a hierarchic method for DTM generation that is based on morphological grey scale opening using structural elements of different sizes (Rottensteiner et al., 2003). Along with cues derived from the other input data, the DTM provides one of the inputs for the second stage, the classification of these data by Dempster-Shafer fusion and the detection of buildings. Five data sets contribute to a Dempster- Shafer fusion process carried out independently for each pixel of the image containing the classification results. After that, initial building regions are instantiated as connected components of building pixels, and a second fusion process is carried out on a per-building level to eliminate regions still corresponding to trees.

2.1 Theory of Dempster-Shafer Fusion

This outline of the theory of Dempster-Shafer is based on (Klein, 1999). We consider a classification problem where the input data are to be classified into n mutually exclusive classes $C_j \in \theta$. The power set of θ is denoted by 2^θ. A probability mass $m(A)$ is assigned to every class $A \in 2^\theta$ by a "sensor" (a classification cue) such that $m(\varnothing) = 0$, $0 \le m(A) \le 1$, and $\Sigma m(A) = 1$, where the sum is to be taken over all $A \in 2^\theta$ and $\varnothing$ denotes the empty set. Imprecision of knowledge can be handled by assigning a non-zero probability mass to the union of two or more classes C_j. The *support Sup(A)* of a class $A \in 2^\theta$ is the sum of all masses assigned to that class:

$$Sup(A) = \sum_{B \subseteq A} m(B) \tag{1}$$

$Sup(\overline{A})$ is the support for the complementary hypothesis of A: $A \cap \overline{A} = \theta$. $Sup(\overline{A})$ represents the degree to which the evidence contradicts a proposition, and it is called *dubiety*. If p sensors are available, probability masses $m_i(B_j)$ have to be defined for all these sensors i with $1 \le i \le p$ and $B_j \in 2^\theta$. The

Dempster-Shafer theory allows the combination of the probability masses from several sensors to compute a combined probability mass for each class $A \in 2^{\theta}$:

$$m(A) = \frac{\sum_{B_1 \cap B_2 \cap \ldots \cap B_p = A} \left[\prod_{1 \le i \le p} m_i(B_j) \right]}{1 - \sum_{B_1 \cap B_2 \cap \ldots \cap B_p = \emptyset} \left[\prod_{1 \le i \le p} m_i(B_j) \right]} \qquad (2)$$

As soon as the combined probability masses *m(A)* have been determined from the original ones, both *Sup(A)* and *Sup($\overline{A}$)* can be computed. The accepted hypothesis $C_a \in \theta$ is the class obtaining maximum support.

2.2 Initial Land Cover Classification

In this process, we want to achieve a per-pixel classification of the input data into one of four classes: buildings *(B)*, trees *(T)*, grass land *(G)*, and bare soil *(S)*. Five cues are used for this purpose, two of them being surface roughness parameters derived by applying polymorphic feature extraction (Förstner, 1994) to the first derivatives of the DSM. They are as follows:

1. The *height differences* ΔH between the DSM and the DTM distinguish elevated objects (trees and buildings) from others. We assign a probability mass $P_{\Delta H} = P_{\Delta H}(\Delta H)$ ascending with ΔH to the combined class $B \cup T$, and $(1 - P_{\Delta H})$ to $G \cup S$.
2. The *height differences* ΔH_{FL} between the first and the last pulse DSMs are large in areas covered by trees. We assign a probability mass $P_{FL} = P_{FL}(\Delta H_{FL})$ ascending with ΔH_{FL} to class T, and $(1 - P_{FL})$ to $B \cup G \cup S$. By doing so we neglect that large values of ΔH_{FL} might also occur at the borders of buildings and at power lines.
3. The *NDVI* is an indicator for vegetation, thus for classes T and G. We assign a probability mass $P_N = P_N(NDVI)$ ascending with *NDVI* to the combined class $T \cup G$, and $(1 - P_N)$ to $B \cup S$.
4. The *strength R of surface roughness*, i.e. the texture strength of polymorphic feature extraction, is large in areas of great variations of the surface normal vectors, which is typical for trees. We assign a probability mass $P_R = P_R(R)$ ascending with R to class T, and $(1 - P_R)$ to $B \cup G \cup S$. By doing so we neglect that large values of R might also occur at the borders of buildings and at step edges of the terrain.
5. The *directedness D of surface roughness*, i.e. the texture directedness of polymorphic feature extraction, is also an indicator for trees, but only if R differs from 0 significantly; otherwise, D is dominated by noise. We assign a probability mass $P_D = P_D(R, D)$ ascending with D to class T, and $(1 - P_D)$ to $B \cup G \cup S$.

The probability masses $P_{\Delta H}$, P_{FL}, P_N, P_R, and P_D are assumed to be equal to a constant P_1 for input parameters $x < x_1$. For input parameters $x > x_2$, they are assumed to be equal to another constant P_2, with $0 \le P_1 < P_2 \le 1$. Between x_1 and x_2, the probability mass is assumed to be a cubic parabola:

$$P_i(x) = P_1 + (P_2 - P_1) \cdot \left[3 \cdot \left(\frac{x - x_1}{x_2 - x_1} \right)^2 - 2 \cdot \left(\frac{x - x_1}{x_2 - x_1} \right)^3 \right] \qquad (3)$$

In equation 3, $i \in \{\Delta H, R, FL, N, D\}$. P_1 and P_2 are chosen to be *5%* and *95%*, respectively. Further, we choose $(x_1, x_2) =$ *(1.5 m, 3.0 m)* for ΔH and ΔH_{FL} and $(x_1, x_2) = (30\%, 65\%)$ for the *NDVI*. With respect to P_R, (x_1, x_2) are linked to the median of R to make the definition of P_R adaptive to the slope variations in a scene: $(x_1, x_2) = [2 \cdot median(R), 15 \cdot median(R)]$. P_D is modelled in the same way with $(x_1, x_2) = (0.1, 0.9)$ if $R < R_{min}$, and by $P_D = 0.5$ otherwise: if the slope variations are not significant, D thus cannot be used to distinguish any of the classes. We choose $R_{min} = 5 \cdot median(R)$. The combined probability masses are computed for each pixel using equation 2, and the pixel is assigned to the class of maximum support. It is an important property of this method that no sharp thresholds are required, but the probability mass functions have a smooth transition between two levels P_1 and P_2.

2.3 Final Classification of Building Regions

After the initial classification, we obtain a binary image of building pixels. Only a small local neighbourhood contributed to the classification of each pixel (via R and D), which causes classification errors, e.g. singular "building" pixels, or "tree" pixels inside buildings. We use a morphological opening filter to eliminate singular building pixels. After that, we create a building label image by a connected component analysis. A second classification based on the Dempster-Shafer theory is applied to the initial building regions thus detected, using four cues representing average values for each building region. The *average height differences* ΔH_a between the DSM and the DTM and the *average NDVI* ($NDVI_a$) are used in the same way as ΔH and *NDVI* in the initial classification. We use different parameters related to surface roughness. The percentage H of pixels classified as "homogeneous" in polymorphic feature extraction is an indicator for an object consisting of smooth surface patches. Thus, we assign a probability mass $P_H = P_H(H)$ to class $B \cup G \cup S$, and $(1 - P_H)$ to T. The percentage P of pixels classified as "point-like" in polymorphic feature extraction is an indicator for trees. We assign a probability mass $P_P = P_P(P)$ to class T, and $(1 - P_P)$ to $B \cup G \cup S$.

The mathematical model described in section 2.2 is also used for computing the probability masses for ΔH_a, H, P, and $NDVI_a$. Again, we choose $P_1 = 5\%$ and $P_2 = 95\%$, further $(x_1, x_2) =$ *(1.5 m, 3.0 m)* for ΔH_a, $(x_1, x_2) = (30\%, 65\%)$ for $NDVI_a$, $(x_1, x_2) = (0\%, 60\%)$ for H, and $(x_1, x_2) = (30\%, 75\%)$ for P. The combined probability masses are evaluated for each initial building region, and if such a region is assigned to another class than "building", it is eliminated. Finally, the building regions are slightly grown to correct for building boundaries erroneously classified as trees.

2.4 Results of Building Detection

Figure 1 shows the results of the initial Dempster-Shafer classification on the left and the final building label image on the right. After morphological opening of the binary image of the building pixels and after eliminating candidate regions smaller than 10 m^2, the second Dempster-Shafer classification is carried out for altogether 2291 building candidate regions. 344 of these regions are found to belong to a class other than "building", so that we finally obtain 1947 building regions.

In the initial classification, class "bare soil" mainly corresponds to streets and parking lots. Trees are situated near the river crossing the scene, along the streets in the residential areas, and in backyards. Step edges at the building boundaries are often classified as trees. Given the resolution of the LIDAR data, it

was not easy to separate trees from buildings in the residential areas. A few residential buildings were erroneously classified as trees, especially if the roof consisted of many small faces. Problems also occurred with bridges, chimneys or other objects on top of large buildings, with parked cars, and with power lines. Shadows in the colour orthophoto were an error source. There are no shadows in the LIDAR intensity data, so that the "pseudo-NDVI" was systematically wrong in these areas.

Figure 1. Left: results of the initial classification. White: grass land. Light grey: bare soil. Dark grey: trees. Black: buildings. Right: the final building label image.

In order to evaluate our method, the *completeness* and the *correctness* (Heipke, 1997) of the results were determined both on a per-pixel and on a per-building level. The evaluation on a per-pixel level shows that 94% of the building pixels were actually detected. The missed buildings were small residential buildings, some having roofs with high reflectance in the wavelength of the laser scanner (thus, a high pseudo-NDVI), others having roofs consisting of many small planar faces, or they are too small to be detected given the resolution of the LIDAR data. For a few larger industrial buildings, some building parts could not be detected due to errors in DTM generation. 85% of the pixels classified as building pixels do actually correspond to a building. This number is affected by errors at the building boundaries, and there are a few larger false positives at bridges, at small terrain structures not covered by vegetation, and at container parks.

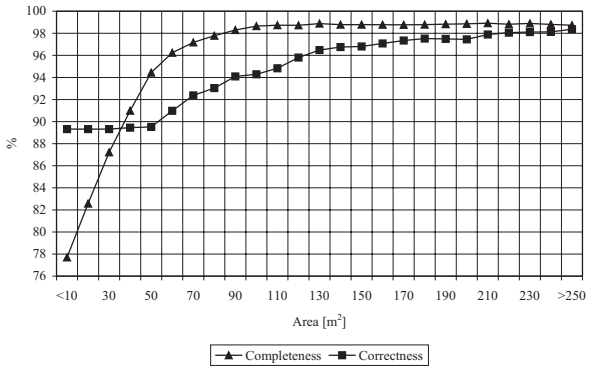

Figure 2. Completeness and correctness of the detection results in dependence of the building size.

The results of the evaluation on a per-building basis are presented in figure 2. It shows the cumulative completeness and correctness for buildings being larger than the area shown in the abscissa. Our algorithm detected 95% of all buildings larger than 50m^2 and 90% of the buildings larger than 30 m^2. Buildings smaller than 30 m^2 (mostly garden sheds or garages) could not usually be detected. The correctness was 96% for buildings larger than 120 m^2 and 89% for all detected regions.

3. FUSION OF LIDAR DATA AND IMAGES FOR ROOF PLANE DETECTION AND DELINEATION

The work flow for the geometric reconstruction of the buildings consists of four steps (Rottensteiner and Briese, 2003):

1. **Detection of roof planes** based on a segmentation of the DSM and/or the image data to find planar segments which are expanded by region growing algorithms.
2. **Grouping of roof planes and model generation:** Co-planar roof segments are merged, and hypotheses for intersection lines and/or step edges are created based on an analysis of the neighbourhood relations. This results in a model consisting of a conglomerate of roof planes, complemented by walls.
3. **Consistent estimation of the model parameters:** The parameters of the building models are improved by a consistent estimation procedure using all the available data.
4. **Model regularisation:** The models are improved by introducing hypotheses about geometric constraints between planes, and parameter estimation is repeated.

In this section we want to show how the fusion of a LIDAR DSM and digital aerial images contributes to an improved detection of planar segments and an improved delineation of the roof boundary polygons. Examples will be presented for the building in figure 3.

Figure 3. Left: DSM of a building (grid width: 0.5 m). Right: aerial image (ground resolution: 0.17 m). Length of the larger wing of the building: 30 m.

3.1 Data Fusion for Roof Plane Detection

The left part of figure 4 shows the planar segments that were extracted from the DSM in figure 3 using the iterative segmentation scheme by Rottensteiner and Briese (2003). The basic structure of the building has been captured, but the segment outlines are very irregular. A proper determination of the roof plane boundaries from these results is difficult for two reasons. First, the segmentation errors cause errors in the neighbourhood relations between the segments, the latter being important prerequisites for checking whether the intersection line between two neighbouring planes is a part of the boundary polygons. Second, the geometric quality of step edges is poor in LIDAR data, and in order to improve it, better approximations are required. In other cases than the one depicted in figure 3, some of the roof planes might actually be missing (Rottensteiner and Briese, 2003). The results of roof plane segmentation in figure 4 can be improved by matching the planar segments detected in the DSM with image segments.

We extract homogeneous segments from the aerial images using polymorphic feature extraction (Foerstner, 1994). In order to mitigate the problem of erroneously merged regions, this is done iteratively, in a similar way as for DSMs (Rottensteiner and Briese, 2003). We use the DSM for geo-coding the results of image segmentation, yielding a label image in object space for each of the aerial images involved (figure 4). The resolution of these label images is chosen in accordance with the image

resolution. In spite of errors in geo-coding caused by errors in the DSM at the building outlines, the segment boundaries in figure 4 match the actual roof boundaries quite well.

Figure 4. Left: planar segments detected in the DSM (fig. 3). Centre/right: Homogeneous segments from two aerial images projected to the DSM. Resolution: 0.1 m.

The DSM label image is resampled to the same resolution as the projections of the image segmentation results. We obtain altogether $N + 1$ label images, where N is the number of aerial images. For each pixel i, we obtain a tuple $h_i = \{l_{Di}, l_{1i}, ..., l_{Ni}\}$ of corresponding labels, indicating a matching candidate between a planar segment l_D and N image segments $\{l_{1i}, ..., l_{Ni}\}$. We determine the number n_i of occurrences for each candidate h_i. We also compute the percentage $p_{ji} = n_i / n_j$ of each segment l_{ji} that contributes to h_i, where n_j is the number of pixels of l_{ji} in the label image j, with $j \in \{D, 1, ..., N\}$.

Due to segmentation and geo-coding errors, the set of matching candidates h_i will contain errors. Hence, we firstly classify each hypothesis h_i according to the percentages p_{ji}. A hypothesis is classified to have *strong* support if it has at least one component j with $p_{ji} > 50\%$. Otherwise, it is said to have *partial* support if there is at least one component with $33\% < p_{ji} \leq 50\%$, or *weak* support if there is there is at least one component with $5\% < p_{ji} \leq 33\%$. If $p_{ji} \leq 5\%$ for all components of the hypothesis h_i or if the number n_i of pixels giving support to it is below a certain threshold, h_i is eliminated. Further, if for a hypothesis h_i there exists another hypothesis $h_k = \{l_{Dk}, l_{1i}, ..., l_{Ni}\}$ that has a higher support than h_i (thus, if there is a planar segment $D_k \neq 0$ corresponding to a larger portion of the co-occurrence of image segments $l_{1i}, ... , l_{Ni}$ than D_i), then h_i is eliminated as well. In this way, contradicting hypotheses between planar segments and tuples of image segments are eliminated.

A set of hypotheses $h_{ij} \in \{h_i\}$ for each planar segment $D_j \neq 0$ is thus obtained, consisting of all hypotheses h_{ij} for which the first component is D_j. We improve the initial segmentation by region growing, taking into account the matching results. Each pixel not yet assigned to a planar segment is tested according to whether it belongs to segment D_j by computing its height from the planar equation of that segment (to avoid errors at the building outlines). The resulting 3D point is back-projected to all images, and the image labels at the projected positions are evaluated. If the set of labels corresponds to one of the hypotheses h_{ij}, the pixel is assigned to segment D_j (figure 5).

In order to improve the initial segmentation by extracting new planar segments, it would be necessary to evaluate the hypotheses h_{i0}, i.e. the hypotheses of multiple image labels corresponding to areas not yet classified. It would be necessary to first compute the plane parameters of these new segments using the DSM. This has not yet been implemented.

A major advantage of our technique is that using multiple images can mitigate segmentation errors. For instance, if two planes are merged in one image due to poor contrast, but correctly separated in another, the algorithm overcomes this problem because any coincidence of two or more image labels is considered to be a new "combined" image label.

The left part of figure 5 shows the results of region growing. The planar segments resemble the roof planes much better than the initial segments from the DSM. The segment boundaries are smoothed by morphologic filtering, and co-planar neighbouring segments are merged, which results in the label image in the centre of figure 5. The right part of figure 5 shows the Voronoi diagram of that label image (Fritsch and Ameri, 2000). It is used to derive the neighbourhood relations of the planar segments, and the boundaries of the segments in the Voronoi diagram provide the first estimates of the roof boundaries.

Figure 5. Left: planar segments after region growing. Centre: the segments after morphologic filtering and merging of co-planar segments. Right: Voronoi diagram.

3.2 Delineation of the 3D Roof Planes

Rottensteiner and Briese (2003) describe how each portion of the original boundary polygon which separates two planes is classified according to whether it is an intersection line, a step edge, or both, in which case that polygon portion is split into smaller parts. The positions of the intersection lines can be determined precisely by the intersection of the neighbouring planar segments. The location of the step edges is critical, especially if it has to be carried out using LIDAR data alone. Here we want to show how the initial boundary polygons can be improved using edges extracted from the aerial images by polymorphic feature extraction.

We start by computing 3D straight line segments from image edges. As the boundary polygon of a roof plane has to be situated in that roof plane, it is possible to match edges extracted from different images by using the assumption that these edges are situated in or at the border of the plane. Thus, in order to delineate the boundary of a roof plane, we project all the image edges in a certain neighbourhood of the approximate polygon to that plane. If the projections of two edge segments l_1 and l_2 from two different images are found to be almost parallel (indicated by a small angle α between their normal vectors, e.g. $\alpha < 15°$) and if there is at least one point on one of the segments that has a distance from the other segment smaller than another threshold (e.g. 0.25 m), the image edges are assumed to be the images of the same straight line in object space. This straight line is computed by adjustment through the end points of the projected image edge segments. The end points of the combined 3D segment are determined so that the combined segment merges both projected image segments (Figure 6). If two such adjusted straight line segments are found to overlap in object space, they will be merged. Thus, we favour long object edges.

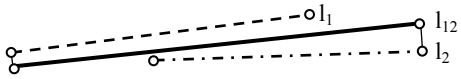

Figure 6. Matching and merging edge segments. l_1 and l_2: edges from images 1 and 2. l_{12}: combined edge, covering the projections of both l_1 and l_2.

Next, these 3D straight line segments have to be matched with the approximate roof polygons. Again, this matching is based on geometric proximity and parallelism, with relatively loose thresholds because of the poor quality of the approximate

polygon. Multiple matches are eliminated based on a minimum distance criterion. The results for the building in figure 3 are presented in figures 7 and 8. The shapes of the boundary polygons could be improved considerably. However, the polygons are not yet completely correct at the building corners. This could be overcome either by instantiating hypotheses about regularities in areas that do not receive support from image features, or by iterating the matching technique in cases where the approximations are not good enough. There is a small displacement between the intersection lines and image edges, which is either caused by errors in the geo-coding or in the plane parameters. This emphasises the importance of a final adjustment, taking into account both the LIDAR points, the image edges (both for step edges and intersection lines), and geometric constraints, using the adjustment model described in (Rottensteiner and Briese, 2003) The considerable improvement of the shapes of the building outlines as compared to figure 5 also improves the prospects for the success of such an adjustment.

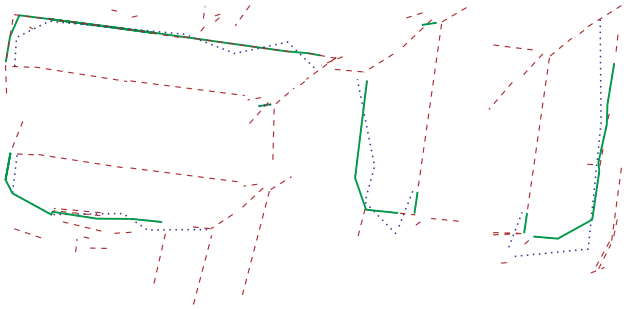

Figure 7. Delineation of step edges for four roof planes. Dashed lines: 3D edges. Dotted line: approximate polygon. Full line: polygon after matching.

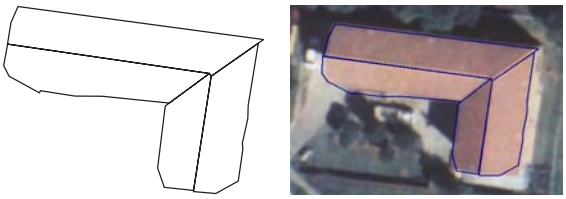

Figure 8. Left: roof polygons after matching. Right: back-projected to one of the aerial images.

4. CONCLUSION AND FUTURE WORK

We have presented a method for building detection from LIDAR data and multi-spectral images, and we have shown its applicability in a test site of heterogeneous building shapes. The method is based on the application of the Dempster-Shafer theory for classification. The results achieved were very satisfactory. The detection rate for buildings larger than 50 m^2 was 95%, and about 89% of the detected buildings were correct. The detection rates decrease considerably with the building size: building structures smaller than 30 m^2 could generally not be detected. In this context, future work will concentrate on evaluating the relative contribution of the cues used for classification. We also want to extend the evaluation to the influence of the LIDAR resolution on the results.

We have also shown how aerial images and LIDAR DSM can be combined to improve both the results of roof plane detection and the shapes of the roof boundaries. This is still work in progress, and the algorithms involved can be improved in many ways. For instance, moments or other invariants of the image segments could be considered in matching, especially if new planar segments are to be introduced based on evidence from the images. The matching of 3D straight lines and roof polygon segments could be expanded to include more robust techniques for outlier detection. However, we have already shown some of the benefits that can be achieved by using multiple data sources for the reconstruction of buildings by polyhedral models.

ACKNOWLEDGEMENTS

This work was supported by the Australian Research Council (ARC) under Discovery Project DP0344678 and Linkage Project LP0230563. The LIDAR data were provided by AAM Hatch, Mt Colah, NSW 2079, Australia (www.aamhatch.com.au). The aerial images are courtesy of AAM Hatch and Sinclair Knight Merz.

REFERENCES

Ameri, B., Fritsch, D., 2000. Automatic 3D building reconstruction using plane roof structures. In: ASPRS congress, Washington, DC, accessed in the WWW (27/01/2004): www.ifp.uni-stuttgart.de/publications/2000/pub2000.html.

Brunn, A., Weidner, U., 1997. Extracting buildings from digital surface models. In: *IAPRS* XXXII / 3-4W2, pp. 27-34.

Förstner, W., 1994. A framework for low level feature extraction. In: *Computer Vision - ECCV '94* Vol. II, 5th ICCV, Boston, MA, pp. 383-394.

Haala, N., Brenner, C., 1999. Extraction of buildings and trees in urban environments. *ISPRS J. Ph & RS* 54(2-3), pp. 130-137.

Heipke, C., Mayer, H., Wiedemann, C., Jamet, O., 1997. Evaluation of automatic road extraction. In: *IAPRS* XXXII / 3-2W3, pp. 47–56.

Klein, L., 1999. Sensor and data fusion, concepts and applications. 2nd edition, SPIE Optical Engineering Press.

Lu, Y.-H., Trinder, J., 2003. Data fusion applied to automatic building extraction in 3D reconstruction. In: *Annual ASPRS conference*, Anchorage, Alaska, pp. 114-122.

Rottensteiner, F., Briese, C., 2003. Automatic generation of building models from LIDAR data and the integration of aerial images. In: *IAPRSIS* XXXIV / 3-W13, pp. 174-180.

Rottensteiner, F., Trinder, J., Clode, S., Kubik, K., 2003. Detecting buildings and roof segments by combining LIDAR data and multispectral images. In: *Image and Vision Computing New Zealand (IVCNZ 2003)*, pp. 60-65.

Schenk, T., Csatho, B., 2002. Fusion of LIDAR data and aerial imagery for a more complete surface description. In: *IAPRSIS* XXXIV / 3A, pp. 310-317.

Voegtle, T., Steinle, E., 2003. On the quality of object classification and automated building modeling based on laserscanning data. In: *IAPRSIS* XXXIV / 3W13, pp. 149-155.

Vosselman, G., Dijkman, S., 2001. 3D building model reconstruction from point clouds and ground plans. In: *IAPRSIS* XXXIV / 3W4, pp. 37-43.