EVALUATION OF A METHOD FOR FUSING LIDAR DATA AND MULTISPECTRAL IMAGES FOR BUILDING DETECTION

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

[a] Institute of Photogrammetry and Remote Sensing, Vienna University of Technology, Gußhausstraße 27-29, 1040 Vienna, Austria - fr@ipf.tuwien.ac.at
[b] School of Surveying and Spatial Information System, The University of New South Wales, UNSW SYDNEY NSW 2052, Australia – j.trinder@unsw.edu.au
[c] 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

Commission III, WG III/4

KEY WORDS: Building detection, Airborne Laser Scanning, Data fusion, Classification, Evaluation

ABSTRACT:

In this paper, we describe the evaluation of a method for building detection by the Dempster-Shafer fusion of LIDAR data and multispectral images. For that purpose, ground truth was digitised for two test sites with quite different characteristics. Using these data sets, the heuristic model for the probability mass assignments of the method is validated, and rules for the tuning of the parameters of this model are discussed. Further we evaluate the contributions of the individual cues used in the classification process to the quality of the classification results. Our results show the degree to which the overall correctness of the results can be improved by fusing LIDAR data with multispectral images.

1. INTRODUCTION

1.1 Motivation and Goals

The high potential of LIDAR data for automatic building detection has been shown in the past, e.g. (Rottensteiner et al., 2004). Building detection essentially is a classification of the input data to separate points situated on buildings from those on other objects. The main problem in this task is to distinguish buildings from trees. Surface roughness parameters can be determined from LIDAR data, but as the resolution of the LIDAR data decreases, the classification becomes more critical in areas where the appearance of trees and buildings in the data is similar. To assist in overcoming this problem, the height differences between the first and the last echoes of the laser and multi-spectral images can be used as additional data sources to improve the classification results.

In our previous work, we have presented a method for fusing first and last pulse LIDAR and multispectral image data based on the theory of Dempster-Shafer (Rottensteiner et al., 2004). It could be shown that the algorithm delivered very good results for an urban test site. However, several questions remained. How does the method perform with data acquired from sensors having other characteristics? How general are the assumptions taken on the model of the probability masses? How is the tuning of the parameters dependent on the input data? How do the individual cues used in data fusion contribute to the quality of the classification results? How well is the NDVI suited for classification in shadow areas? It is the goal of this paper to give answers to these questions, based on a thorough evaluation of our method using two data sets of quite different character-istics. We will also describe how the method was improved based on the experience gained in the course of this project.

* Corresponding author.

1.2 Related Work

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 (Brunn and Weidner, 1997), or by hierarchic robust linear prediction (Rottensteiner and Briese, 2002). 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 (Rottensteiner et al., 2004).

Various classification techniques have been applied for building detection, e.g., unsupervised classification (Haala and Brenner, 1999), rule-based classification (Rottensteiner and Briese, 2002), Bayesian networks (Brunn and Weidner, 1997), and fuzzy logic (Vögtle and Steinle, 2003; Matikainen et al., 2003). The probabilistic approaches among the cited ones face the difficulty of having to model all a priori probabilities, which is problematic if the assumption of a normal distribution of the data vectors is unrealistic. The theory of Dempster-Shafer can help overcome these problems as its capability to handle 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; Rottensteiner et al., 2004).

Vögtle and Steinle (2003) evaluated their method for building detection using two test data sets of 1 m resolution and achieved detection rates of 93% and 96%, respectively. The authors state that the classification accuracy decreases with the building size, without quantifying this effect. Matikainen et al. (2003) used LIDAR data for building change detection. Their method

detected 90% of all building pixels in a reference map, with a false-alarm rate of 20%. For buildings larger than 200 m^2, completeness and correctness are 91% and 84%, respectively; the respective values for buildings smaller than 200 m^2 are 42.1% and 34.9%. A minimum percentage of overlap of 70% is required for a building to be classified as "correctly identified".

Vosselman et al. (2004) first separate bare earth LIDAR points from other points and then further classify the other points according to whether they belong to buildings or trees. They apply their classification to the original LIDAR point clouds. Their results for points on buildings correspond to a completeness of 85% and a correctness of 92%. In their conclusions they state that using additional colour information increased the classification accuracy for buildings by 3%.

1.3 The Test Data Sets

We use two test data sets, one captured over Fairfield (NSW) using an Optech ALTM 3025 laser scanner, and the other over Memmingen (Germany) with a TopoSys scanner. Both cover an area of 2 x 2 km^2, and both contain the first and the last echoes of the laser beam. The characteristics of the two test areas are quite different. Fairfield covers a suburban area with low density of development in the southwest half of the scene, whereas the northeast part is dominated by large industrial buildings. The trees are mostly Eucalypts and thus are evergreen. Memmingen features a densely developed historic centre in the north of the scene and industrial areas in the remainder. The Memmingen data set was captured at a time when the trees had hardly any foliage, so that a much larger proportion of last pulse points were reflected by the ground. The multi-spectral information was also quite different for the two data sets. For Fairfield, an RGB orthophoto with a resolution of 0.15 m was available (figure 1). We created a "pseudo-NDVI-image" at a resolution of 1 m, using the red band from the orthophoto and substituting the LIDAR intensity values for the infrared band. For Memmingen, geo-coded RGB and CIR (colour infrared) images with a resolution of 0.5 m were available (figure 1). The infrared bands of the data sets thus corresponded to different wavelengths.

From the original LIDAR point clouds, DSM grids were interpolated for both the first and the last pulse echoes by linear prediction using a small degree of smoothing (Rottensteiner and Briese, 2002). The original point distances, the grid widths Δ of the DSM grids, and the wavelengths of the infrared band used for computing the NDVI are shown in table 1.

	S_i	S_c	Δ	IR	n_B	B_p	T_p
F	1.2	1.2	1.00	1047	2424	26	17
M	0.2	1.2	0.75	770-890	2046	17	28

Table 1. Characteristics of the test datasets. F: Fairfield, M: Memmingen. S_i, S_c: point spacing in and across flight direction [m]; Δ: grid width of the interpolated DSM [m]; IR: wavelength of the infrared band [µm]. n_B: number of buildings digitised. B_p, T_p: percentage of area covered by buildings and trees [%].

Reference data were captured by digitising buildings and trees in the orthophotos. We chose to digitize all structures recognisable as buildings or trees independent of their size. Neighbouring buildings that are joined but are obviously separate entities were digitized as separate polygons. Larger areas covered by trees were digitised as one polygon. No information on single trees was captured. The numbers of building polygons and the percentage of the area covered by buildings and trees are shown in table 1. In Fairfield, the orthophoto and the LIDAR data correspond to different epochs. We thus had to exclude 49 building polygons that were only available in one data set from our analysis.

Figure 1. Orthophotos for Fairfield (left), Memmingen (right).

2. THE ALGORITHM FOR BUILDING DETECTION

The input to our method is given by four data sets that have to be generated from the raw data by pre-processing: the two DSM grids corresponding to the first and the last pulse data, the DTM, and the NDVI. The DTM is optional because it could also be derived from the last pulse DSM by hierarchic morphologic filtering (Rottensteiner et al., 2004). As we only want to evaluate our classification method based on Dempster-Shafer fusion, we chose to determine DTM grids by robust linear prediction (Rottensteiner and Briese, 2002).

The work flow for building detection consists of three stages. In the first stage there are five "sensors" (classification cues) that contribute to a Dempster-Shafer fusion process carried out for each pixel of the DSM grid independently. The second stage is a post-classification process that aims at eliminating errors of the initial classification at building outlines. Morphological filtering of the binary "building image" eliminates small areas of pixels erroneously classified as building pixels. Connected components of building pixels are then sought, which results in initial building regions. The third stage of building detection is a Dempster-Shafer fusion process carried out on a per-building level to eliminate spurious initial building regions.

2.1 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 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. The plausibility $Pls(A)$ sums up al probability masses not assigned to the complementary hypothesis $\overline{A}$ of A with $A \cap \overline{A} = \varnothing$ and $A \cup \overline{A} = \theta$:

$$Sup(A) = \sum_{B \subseteq A} m(B)\,; \qquad Pls(A) = \sum_{A \cap B \neq \emptyset} m(B) = 1 - Sup(\overline{A}) \quad (1)$$

*Sup($\overline{A}$)*is called *dubiety*. It represents the degree to which the evidence contradicts a proposition. 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 $B_j \in 2^\theta$. From these probability masses, a combined probability mass can be computed for each class $A \in 2^\theta$:

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

The sum in the denominator of equation 2 is a measure of the *conflict* in the evidence. As soon as the combined probability masses *m(A)* have been determined, both *Sup(A)* and *Pls(A)* can be computed. The accepted hypothesis $C_a \in \theta$ is determined according to a decision rule, e.g. as the class of maximum plausibility or the class of 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 "sensors" are used for this purpose. In our model for the distribution of the evidence from each sensor to the four classes, we assume in general that each sensor i can separate two complementary subsets U_{Ci} and $\overline{U}_{Ci}$ of θ. The probability mass $P_i(x_i)$ assigned to U_{Ci} by the sensor i depending on the sensor output x_i is assumed to be equal to a constant P_1 for $x_i < x_1$. For $x_i > x_2$, it is 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 heuristically assumed to be a cubic parabola with horizontal tangents at $x_i = x_1$ and $x_i = x_2$, yielding a smooth transition between the probability levels P_1 and P_2:

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

The probability mass *[1 - $P_i(x_i)$]* will be assigned to class $\overline{U}_{Ci}$. The combined probability masses are computed for each pixel, and the pixel is assigned to the class of maximum support / plausibility. In the following sections, we will describe this model and its modifications for each of the five "sensors". We will also validate the model using the ground truth and show how the model parameters can be chosen.

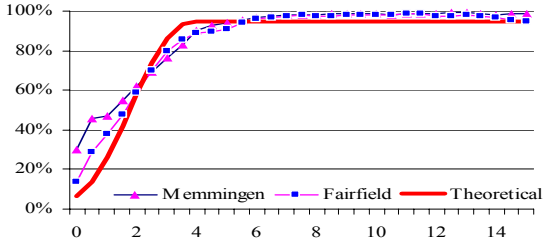

Figure 2: $P_{\Delta H}$ *(ΔH)* for $B \cup T$ from ground truth compared to the theoretical model. Abscissa: *ΔH* [m].

2.2.1 Height differences *ΔH* between DSM and DTM: *ΔH* distinguishes elevated objects from others. We assign a probability mass $P_{\Delta H} = P_{\Delta H}(\Delta H)$ according to the model described above to $B \cup T$, and *(1 - $P_{\Delta H}$)* to $G \cup S$. The last pulse DSM should be used to optimise the classification accuracy for buildings. Figure 2 compares $P_{\Delta H}$ *(ΔH)* derived from the ground truth to the theoretical model with $(P_1, P_2) = (5\%, 95\%)$ and $(x_1, x_2) = (0\ m, 4\ m)$. The theoretical model fits quite well to the data of both Memmingen and Fairfield. The differences for small values of *ΔH* correspond to pixels within the tree polygons having a last pulse return from the terrain. We consider the parameters given above to be generally applicable.

2.2.2 Strength of surface roughness: Surface roughness R, i.e. the texture strength of polymorphic feature extraction (Förstner, 1994) applied to the first derivatives of the DSM, is large in areas of great variations of the surface normal vectors, which is typical for trees. The absolute values of R will vary with the scene, so that it is impossible to find values for the parameters of the model described by equation 3 that are generally applicable. The situation can be improved by a re-parameterisation of R. Rather than using R, we characterise surface roughness by the percentage $R_P(R)$ of pixels for which the surface roughness is smaller than R. $R_P(R)$ is limited to the interval [0..100%]. We assign a probability mass $P_R = P_R(R_P)$ 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.

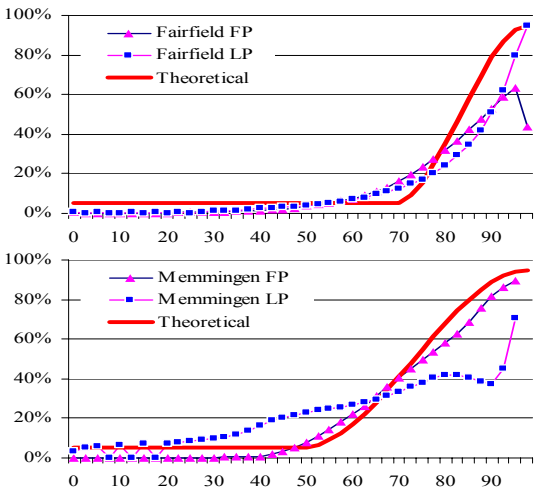

Figure 3: $P_R(R_P)$ for T from ground truth compared to the theoretical model for Fairfield (above) and Memmingen (below). LP/FP: Last / First pulse data. Abscissa: R_P [%].

Assuming that the trees correspond to the areas of maximum surface roughness, we can use an estimate for the percentage of the scene covered by trees to derive the values of the parameters of the model in equation 3. Denoting the percentage of trees by T_p, x_1 can be determined so that $P_R(R_P) = 50\%$. Using $(P_1, P_2) = (5\%, 95\%)$ and $x_2 = 100\%$ yields:

$$x_1 = x_2 - 2 \cdot T_p \qquad (4)$$

Using rounded values for T_p according to table 1, this results in $x_1 = 70\%$ for Fairfield and $x_1 = 50\%$ for Memmingen. Figure 3 compares the distribution of $P_R(R_P)$ from ground truth with the theoretical values thus obtained. It shows that the model fits quite well to the last pulse data in Fairfield, with a larger deviation for first pulse data which is mainly caused by powerlines. With Memmingen, the model fits quite well to the first pulse data, but not so well to the last pulse data. This is caused by the high penetration rate of the laser pulse in this data set. There are many laser strikes on the ground in the forested areas, which means that the DSM almost corresponds to the

(smooth) DTM in these areas. As a consequence, we let the user decide whether to use first or last pulse data for computing surface roughness, depending on which data set represents the roughness of the DSM for vegetation in a better way. In Fairfield, we use last pulse data, in Memmingen, first pulse.

2.2.3 Directedness of surface roughness: The directedness D of surface roughness, i.e. the texture directedness of polymorphic feature extraction (Förstner, 1994), is also an indicator for trees, but only if R differs significantly from 0; otherwise, D is dominated by noise. We assign a probability mass $P_D = P_D(R, D)$ to class T, and $(1- P_D)$ to $B \cup G \cup S$. The same considerations with respect to using first or last pulse data as for R hold true. In order to decide whether R is significant or not, we have to compare it to a threshold R_{min}. We determine R_{min} so that T_p percent of the data have $R > R_{min}$; thus, D will be considered for the T_p "roughest" pixels. We select $(P_1, P_2) = (10\%, 70\%)$ and $(x_1, x_2) = (0, 1)$. The lower value for P_2 should compensate for the fact that large values for D occur with buildings having many small roof planes. Figure 4 shows the distribution of D and the theoretical values, using the same values for T_p as in section 2.2.2. It fits quite well for the last pulse data in Fairfield. There are some deviations in the Memmingen data set, but the model should be applicable there as well. The most important parameter is the estimate for T_p.

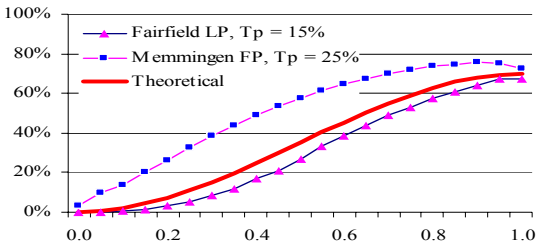

Figure 4: $P_D\ (D)$ for T from ground truth compared to the theoretical model. Abscissa: D.

2.2.4 Height differences ΔH_{FL} between first and last pulse: ΔH_{FL} is large in areas covered by trees. We assign a probability mass $P_{FL} = P_{FL}\ (\Delta H_{FL})$ to class T. However, a small value of ΔH_{FL} does not necessarily mean that there are no trees. That is why we assign the probability mass $(1 - P_{FL})$ to θ and not to $B \cup G \cup S$. We consider the values $(P_1, P_2) = (5\%, 95\%)$ and $(x_1, x_2) = (0\ m, 4\ m)$ to be generally applicable. Figure 5 compares $P_{FL}\ (\Delta H_{FL})$ derived from ground truth to the theoretical model. The empirical curves are not quite as steep as the theoretical one, which is caused by large values of ΔH_{FL} at building outlines. In Fairfield, the upper limit for the probability mass seems to be 80% rather than 95%, which is caused by large values of ΔH_{FL} at powerlines.

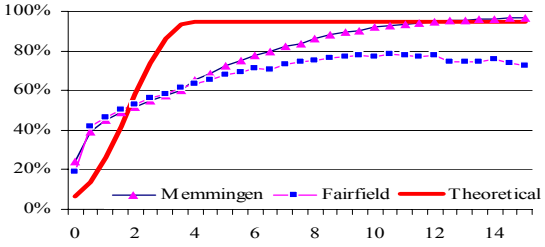

Figure 5: $P_{FL}\ (\Delta H_{FL})$ for T from ground truth compared to the theoretical model. Abscissa: ΔH_{FL} [m].

2.2.5 NDVI: The NDVI is an indicator for vegetation, thus for classes T and G. According to the general model described above, we would assign a probability mass $P_N = P_N\ (NDVI)$ to $T \cup G$, and $(1- P_N)$ to $B \cup S$. This model has to be improved to consider the uncertainty of the NDVI in shadow areas. The NDVI is defined as the ratio between the difference and the sum of the infrared band IR and the red band Rd. Its standard deviation σ_{NDVI} can be computed from:

$$\sigma_{NDVI} = \frac{2 \cdot \sqrt{Rd^2 \cdot \sigma_{IR}^2 + IR^2 \cdot \sigma_{Rd}^2}}{(IR + Rd)^2} \qquad (5)$$

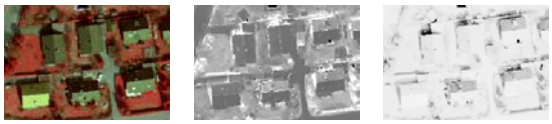

Figure 6: CIR image (left), $NDVI$ (centre) and σ_{NDVI} (right). In the σ_{NDVI} image, white areas correspond to low σ_{NDVI}.

In equation 5, σ_{Rd} and σ_{IR} are the standard deviations of Rd and IR. They are determined by analysing the first derivatives of IR and Rd (Förstner, 1994). Figure 6 shows the CIR image, the NDVI, and σ_{NDVI} for a part of the Memmingen data set. For buildings with a ridge in east-west direction, the NDVI is 0.25 in the shady roof planes and -0.1 in the sunny areas, with standard deviations of about ±0.1 and ±0.02, respectively. The NDVI suggests a strong support for classifying the roof planes in the shade as vegetation. This is compensated for by modulating the probability masses depending on σ_{NDVI}. For $\sigma_{NDVI} \geq 0.25$, we assign a probability mass of 1.0 to θ, i.e., the NDVI will not contribute to the classification. If $\sigma_{NDVI} < 0.25$, $P_{N\theta} = 2 \cdot \sigma_{NDVI}$ is assigned to θ. Using $P_N^0 = P_N\ (NDVI)$ according to equation 3, we assign $P_N = (1 - 2 \cdot \sigma_{NDVI}) \cdot P_N^0$ to class $T \cup G$ and $P_{Ninv} = (1 - 2 \cdot \sigma_{NDVI}) \cdot (1 - P_N^0)$ to $B \cup S$.

We do not have ground truth for the classes G and S, so that we cannot check our model for the NDVI. Given the different spectral characteristics of the infrared bands for the two data sets (cf. table 1) and the fact that the NDVI depends on the lighting conditions and the predominant type of vegetation, we think that the parameters for the model have to be determined in a training phase. Here we choose $(P_1, P_2) = (10\%, 90\%)$ for both data sets. Further, we choose x_1 and x_2 to be $(-0.3, 0.3)$ for Fairfield and $(-0.1, 0.3)$ for Memmingen.

2.3 Post-Classification

We apply a rule-based technique for post-classification in order to eliminate single building pixels and to compensate for errors at the building outlines in our model for the probability masses. Two steps are carried out in an iterative way until a maximum number of iterations is reached. First, we check pixels where the classification is insecure due to contradicting evidence as indicated by a conflict $K > 50\%$. For these pixels, we compute a histogram of the classes in a 5 x 5 neighbourhood. If the class C_{5x5} having the maximum number of occurrences is identical to the class C_2 achieving the second highest score according to the decision rule in the Dempster-Shafer fusion (cf. section 2.1), the classification of the pixel is changed to C_2. In the second step, we determine the class C_{3x3} having a maximum number of occurrences in a 3 x 3 neighbourhood of each pixel. The classification is changed if C_{3x3} is identical to C_2 or if all pixels except the central one are assigned to C_{3x3}. In the resulting binary building image, small elongated areas of building pixels are eliminated by morphologic opening. After that, a building label image is created by a connected component analysis.

2.4 Region-based Classification

A second classification based on the Dempster-Shafer theory is applied to the initial building regions. 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. In computing $NDVI_a$, the individual NDVI values are weighted by $1/\sigma_{NDVI}^2$. We use the percentage P of pixels classified as "point-like" in polymorphic feature extraction as 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 percentage H of pixels classified as "homogeneous" in polymorphic feature extraction (Rottensteiner et al., 2004) is no longer used because it eliminates too many buildings having small roof planes.

The mathematical model described in section 2.2 is also used for computing the probability masses for ΔH_a, P, and $NDVI_a$. For $NDVI_a$ we choose the same model as for the NDVI. For the remaining sensors, we select $(P_1, P_2) = (5\%, 95\%)$. For ΔH_a, the values for $(x_1, x_2) = (1\ m, 3\ m)$ are chosen to be a bit tighter than for ΔH. The parameters for P depend on the average size of a roof plane in relation to the LIDAR resolution. They have to be determined in a training phase. Here, we selected $(x_1, x_2) = (25\%, 75\%)$ for P. No ground truth is available for validating the models for P and $NDVI_a$. The combined probability masses are evaluated for each initial building region, and any region assigned to another class than "building" is eliminated.

3. RESULTS AND EVALUATION

3.1 Evaluation of the Classification Results

Figure 7 shows the results of the Dempster-Shafer classification after post-processing. We used the model for probability masses described in section 2 and eliminated all candidate regions smaller than 10 m^2. Table 2 shows the numbers of buildings detected in the two data sets before and after the region-based classification. Region-based classification eliminates many false candidates in the Fairfield data set. For the Memmingen data set, only a few buildings are actually eliminated at that stage. In the initial classification, class "bare soil" mainly corresponds to streets, parking lots, railway lines, and bare fields. Step edges at the building boundaries are often classified as trees, an effect that is reduced but not completely eliminated by post-classification. Problems occurred with very small buildings, bridges, chimneys or other objects on top of large buildings, with lorries, and with power lines.

In order to evaluate our method, the *completeness* and the *correctness* of the results were determined both on a per-pixel and on a per-building level using the methodology described in (Rottensteiner, et al., 2004). The results of the evaluation on a per-pixel level are also shown in table 2. In both data sets, 90% of the building pixels were detected. The missed buildings were mostly small residential buildings having roofs consisting of many small faces, or they were too small to be detected given the resolution of the LIDAR data. A slightly larger percentage of the pixels classified as building pixels actually correspond to a building. False positives mostly occur at bridges, at small terrain structures not covered by vegetation, and at container parks. Post-classification and region-based classification improve completeness by 5%, but decrease correctness by about half that amount. These numbers have to be interpreted with caution because they are affected by errors in the reference data. The orthophotos were generated using a DTM, so that buildings are shifted away from the nadir point of the sensor. In Fairfield, this can be up to 5 m for a building of 10 m in height.

Figure 7. Results of the Dempster-Shafer classification after post-processing for Fairfield (left) and Memmingen (right). Light green: grass land. Brown: bare soil. Dark green: trees. Red: buildings.

	B_i	B_f	Cp_i / Cr_i [%]	Cp_f / Cr_f [%]
F	2019	1737	84.8 / 96.0	90.2 / 93.3
M	2124	2102	85.0 / 94.2	90.1 / 91.9

Table 2. B_i/B_f: Number of buildings before / after region-based classification. F / M: Fairfield / Memmingen. Cp_i / Cr_i and Cp_f / Cr_f: completeness / correctness for building pixels with and without region-based classification.

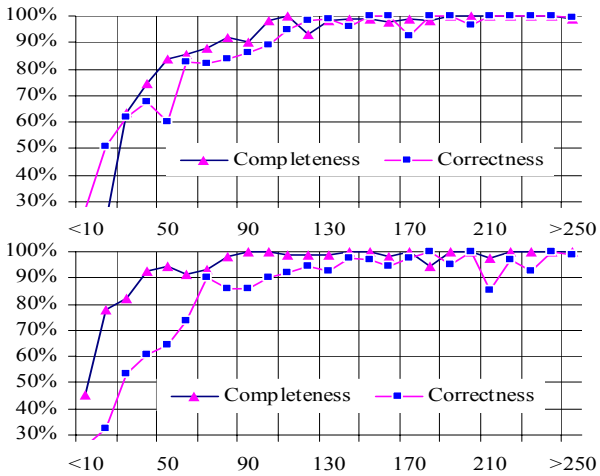

Figure 8. Completeness and correctness of the detection results as a function of the building size [m^2] for Fairfield (above) and Memmingen (below).

For the evaluation on a per-building basis, a building in the reference data set is counted as a true positive if at least 50 % of its area is covered by buildings detected automatically and vice versa (Rottensteiner et al., 2004). Figure 8 shows the completeness and correctness for buildings depending on the building size. In Fairfield, buildings larger than 100 m^2 can be detected reliably, with both completeness and correctness being larger than 90%. Our algorithm could detect 80% of the buildings with an area between 50 m^2 and 100 m^2, and the majority of buildings with an area between 30 m^2 and 50 m^2 could still be detected. Buildings smaller than 30 m^2 were not detectable. With Memmingen, a greater percentage of smaller buildings could be detected, at the cost of a somewhat higher ratio of false positives with an area of 30 m^2 to 50 m^2. False positives are eliminated by region-based classification in Fairfield, but not in Memmingen. Looking at the cumulative completeness / correctness, we can say that 95% of *all*

buildings larger than 40 m² (Fairfield) and *20 m²* (Memmingen) could be detected, whereas 95% of *all detected buildings larger than 50 m²* (Fairfield) and *70 m²* (Memmingen) were correct.

3.2 Contributions of the Individual Classification Cues

In order to evaluate the contributions of the individual sensors to the accuracy of building detection, we apply our method for building detection to four different combinations of sensors. In the variant *LP*, we only use last pulse LIDAR data, thus only the three cues *ΔH, R,* and *D* are used. In Memmingen, the model for surface roughness has to be updated using T_p = 15% for the percentage of trees visible in the last pulse data. In variant *LP+FP*, we combine first and last pulse data. In variant *LP+NDVI*, we use last pulse data and the NDVI. The last variant uses all available data as described in section 3.1.

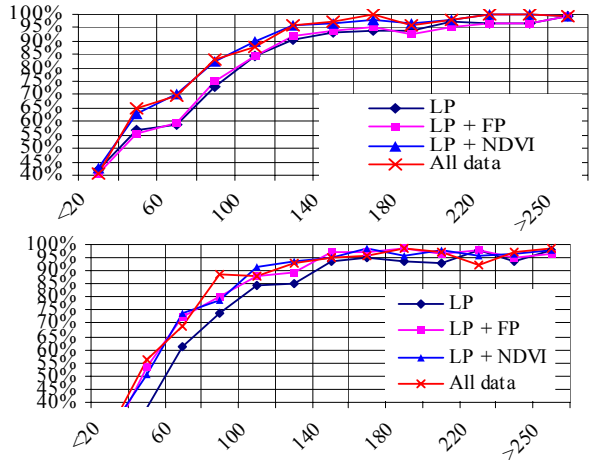

Figure 9. Correctness of the detection results in dependence of the building size [m²] for the four variants described in the text [%]. Above: Fairfield, below: Memmingen.

The completeness achieved in all these variants is almost identical for the Fairfield data set. In Memmingen, variant *LP* performs worse than the others for buildings smaller than 150 m², for which 15% of the buildings between 90 m² and 120 m² are missing. The other variants differ by 1% - 2% for buildings smaller than 40 m², with variant *LP + NDVI* generally performing best. The correctness shows larger differences between the variants. In Fairfield, variants *LP* and *LP + FP* perform equally well. Adding the NDVI to *LP* increases correctness by about 3% for buildings larger than 100 m², but by about 10% for buildings smaller than that limit. As the limits of classification by LIDAR data are reached, the NDVI substantially increases correctness. In Memmingen, variant *LP* performs worse than the others. Variants *LP + FP* and *LP + NDVI* perform almost equally well. Adding the first pulse or the NDVI to *LP* increases the correctness by 5% - 15%, with the increase growing with decreasing building size. The order of all variants except *LP* is not so obvious from figure 9. As a tendency, the NDVI will increase correctness by about 2% - 4% compared to variant *LP + FP*, though in some area intervals, variant *LP + FP* performs slightly better. Variant *LP + NDVI* performs equally well as the variant using all data.

4. CONCLUSION

We have presented a method for building detection based on Dempster-Shafer fusion of LIDAR data multispectral images. We have validated the assumptions of the model for assigning probability masses using two data sets comprising both different sensor and scene characteristics. For the pixel-based classification we found simple rules for setting the parameters of that model if an estimate for the area covered by trees is known. Only the parameters for the NDVI seem to require training. We have improved the method by considering the uncertainty of the NDVI and by post-classification. We have also evaluated the method, giving detailed quality measures on a per-pixel and on a per-building basis. Buildings larger than 50 m² could be reliably detected in both data sets. An investigation into the contributions of the individual cues showed that the NDVI increases the correctness by up to 15% for smaller buildings. First pulse data also help, though to a lesser degree.

In the future, we want to improve the region-based classification method by investigating other cues than those currently in use. Our method performed well in Fairfield, but not so well in Memmingen. We further want to evaluate the sensitivity of the results with respect to the parameter settings, and the relation between completeness / correctness of the results and the resolution of the LIDAR data.

ACKNOWLEDGEMENTS

This work was supported by the Australian Research Council (ARC) under Discovery Project DP0344678 and Linkage Project LP0230563. The Fairfield data set was provided by AAM Hatch (www.aamhatch.com.au). The Memmingen data set was provided by TopoSys (www.toposys.com).

REFERENCES

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.

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

Matikainen, L., Hyyppä, J., Hyyppä, H., 2003. Automatic detection of buildings from laser scanner data for map updating. In: *IAPRSIS* XXXIV / 3W13, pp. 218-224.

Rottensteiner, F., Briese, C., 2002. A new method for building extraction in urban areas from high-resolution LIDAR data. In: *IAPRSIS*, XXXIV / 3A, Graz, Austria, pp. 295 – 301.

Rottensteiner, F., Trinder, J., Clode, S., Kubik, K., 2004. Using the Dempster Shafer method for the fusion of LIDAR data and multi-spectral images for building detection. *Information Fusion* 6(4), pp. 283-300.

Vögtle, 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., Gorte, B., Sithole G., 2004. Change detection for updating medium scale maps using laser Altimetry. In: *IAPRSIS* XXXV - B3, pp. 207-212.