Building Detection by Dempster-Shafer Fusion of LIDAR Data and Multispectral Aerial Imagery

Franz Rottensteiner, John Trinder
School of Surveying and SIS
The University of New South Wales
Sydney, NSW 2052, Australia
{f.rottensteiner, j.trinder}@unsw.edu.au

Simon Clode, Kurt Kubik, Brian Lovell
Intelligent Real-Time Imaging and Sensing Group
School of ITEE, University of Queensland
Brisbane, QLD 4072, Australia
{sclode, kubik, lovell}@itee.uq.edu.au

Abstract

A method for the classification of land cover in urban areas by the fusion of first and last pulse LIDAR data and multi-spectral images is presented. Apart from buildings, the classes "tree", "grass land", and "bare soil" are also distinguished by a classification method based on the theory of Dempster - Shafer for data fusion. Examples are given for a test site in Germany.

1. Introduction

The high potential of LIDAR (*LI*ght *D*etection *A*nd *R*anging) data for the automatic generation of 3D city models has been shown in the past [6]. LIDAR delivers a digital surface model (DSM), i.e. a model of the earth as seen from the sensor's vantage point, including objects on the terrain. The first step of building extraction from LIDAR data is to separate terrain points from non-terrain points thus creating a digital terrain model (DTM), a model of the terrain only. Points on buildings are then separated from points on trees by evaluating the surface roughness [6], [9]. With decreasing resolution of the LIDAR data, the classification becomes more difficult 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 multispectral images can be used to improve the classification results.

In [1], the classification of a LIDAR DSM is accomplished by a hierarchic framework for combining various cues derived from DSM in a Bayesian network. However, estimating the conditional probabilities required for the Bayesian network is very complex. In [3] a normalised DSM created by morphologic filtering is used as an additional band along with colour infrared (CIR) imagery in an unsupervised classification algorithm. The clusters in feature space are assigned to thematic classes interactively, so that the interpretation itself is not automated.

It is the goal of this paper to give a contribution to the fusion of first and last pulse LIDAR data and multi-spectral images for an improved classification of land cover in urban areas, with an emphasis on the detection of buildings. This is accomplished by a method applying the theory of Dempster-Shafer for data fusion [8], [4]. Examples are presented for a test site in Memmingen (Germany).

2. Method Overview

The input to our method is given by three data sets: **(1)** The *last pulse DSM* is sampled into a regular grid by linear prediction with a low degree of filtering [6]. **(2)** The first pulse DSM is also sampled into a regular grid, and by computing the height differences between these DSMs, we get a model of the *height differences between the first and the last pulses* ΔH_{FL}. **(3)** The *normalised difference vegetation index (NDVI)* is computed from the near infrared and the red bands of a geocoded multi-spectral image [5]. The work flow for building detection consists of two stages. First, a coarse DTM is generated. 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.

DTM generation: We use a hierarchic method for DTM generation that is based on morphological grey scale opening using structural elements of different sizes. After morphological opening, a rule-based classification algorithm is used to detect large buildings. In the next iteration, morphological opening is applied using a smaller structural element, but in the building regions already detected, the DTM heights of the previous iteration are substituted for the results of morphological filtering. The process is finished as soon as the minimum size for the structural element is reached. This method is described in detail in [7].

Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)
1051-4651/04 $ 20.00 IEEE

Building Detection Using Dempster-Shafer Fusion: There are altogether five data sets that contribute to a Dempster-Shafer fusion process carried out for each pixel of the image containing the classification results independently: the height differences ΔH between the DSM and the DTM, ΔH_{FL}, the NDVI, and two surface roughness parameters. Each pixel is assigned to one of four classes, namely *building*, *tree*, *grass land*, and *bare soil*. In the subsequent steps, the binary image of the pixels classified as buildings is used. Morphologic filtering helps to eliminate single building pixels and oddly shaped areas of building pixels. After that, initial building regions are instantiated as connected components of building pixels. The average NDVI, the average height differences, and two surface roughness parameters are computed for each of the building regions. These attributes are used in a second fusion process to eliminate regions still corresponding to trees. Building detection using Dempster-Shafer fusion is described in section 3.

3. Data Fusion for Building Detection

We start with an outline of Dempster-Shafer fusion based on [4] and [5]. 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. It contains both the *original classes* C_j and all their possible unions (hence called *combined classes*). A probability mass $m(A)$ is assigned to every class $A \in 2^\Theta$ by a "sensor" (a classification cue) such that $m(\emptyset) = 0$, $0 \leq m(A) \leq 1$, and $\sum_{A \in 2^\Theta} m(A) = 1$, where $\emptyset$ 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 A:

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

$Sup(\bar{A})$ is the support for the complementary hypothesis of A: $A \cup \bar{A} = \Theta$. $Sup(\bar{A})$ 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 $1 \leq i \leq p$ 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 $A \in 2^\Theta$:

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

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

3.1. 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 derived from the input data are used for this purpose:

(1) The *height differences* ΔH between the DSM and the DTM distinguish elevated objects, i.e. trees and buildings, from other objects. 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 the occurrence of 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 [2] 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. We assign a probability mass $P_R = P_R(R)$ ascending with R to class T, and $(1 - P_R)$ to $B \cup G \cup S$. 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 [2] applied to the first derivatives of the DSM, is large in areas of great, but isotropic local variations of the surface normal vectors. Again, this is an indicator for trees, but only if R is above a certain threshold; 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_R, P_{FL}, and P_N 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 $0 \leq P_1 < P_2 \leq 1$. Between x_1 and x_2, the probability mass is described by a cubic parabola using $\bar{x} = \frac{x - x_1}{x_2 - x_1}$ and $i \in \{\Delta H, R, FL, N\}$:

$$P_i(\bar{x}) = P_1 + (P_2 - P_1) \cdot \left(3 \cdot \bar{x}^2 - 2 \cdot \bar{x}^3\right) \tag{3}$$

P_1 and P_2 are chosen to be 1% and 99%, respectively. Further, we choose (x_1, x_2) = (1.5 m, 3.0 m) for ΔH and ΔH_{FL} and (x_1, x_2) = (46%, 66%) 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. The probability mass P_D is assumed to be $P_D = P_1 = 1\%$ if $(D < D_{min}) \vee (R < R_{min})$ and $P_D = D$ otherwise, with D_{min} = 0.5 and $R_{min} = 5 \cdot median(R)$.

The combined probabilities are computed for each pixel independently using equation 2, and the pixel is assigned to the class of maximum support. Compared to the method described in [6], there are several improvements. Firstly, all cues are evaluated simultaneously, thus, we do not eliminate one part of the data prior to the others being used for classification. Instead we get an overall classification of land cover with respect to our four classes. Secondly, no sharp thresholds are required, but the probability mass functions have a smooth transition between the levels P_1 and P_2.

3.2. Final Classification of Building Regions

After the initial classification, we obtain a binary image of pixels classified as "building". As all pixels were classified independently, only a small local neighbourhood contributed to the classification (via R and D). This introduces classification errors such as the appearance of singular "building" pixels inside larger areas of other classification or "tree" pixels inside building roofs. In order to eliminate singular building pixels, we use morphologic opening.

A building label image is then created by a connected component analysis. The individual building regions thus detected are classified based on the Dempster-Shafer theory using four cues representing average values for each building region. The first two cues, i.e. 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 initinal classification. Thirdly, the percentage H of pixels classified as "homogeneous" in polymorphic feature extraction [2] 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. Finally, 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 $B \cup G \cup S$.

For the probability masses $P_{\Delta H_a}$, P_H, P_P, and P_{N_a} we use the same function as for $P_{\Delta H}$ (section 3.1). Again, we choose P_1 = 1% and P_2 = 99%, further (x_1, x_2) = (1.5 m, 3.0 m) for ΔH_a, (x_1, x_2) = (46%, 66%) for $NDVI_a$, (x_1, x_2) = (0%, 50%) for H, and (x_1, x_2) = (40%, 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.

4. Experiments

The test data set was captured over Memmingen (Germany) by TopoSys. An area of 2 × 2 km^2 was processed. Both the first and the last pulses were recorded. The average LIDAR point distance was about 1.2 m (cross-flight) by 0.20 m (in-flight), from which we derived DSM grids at a resolution of 0.75 m. A geocoded false-colour infrared image with a resolution of 0.5 m was also available. We computed the NDVI from the infrared and the red bands of that image, respectively (figure 1).

Figure 1. DSM (left) and NDVI (right).

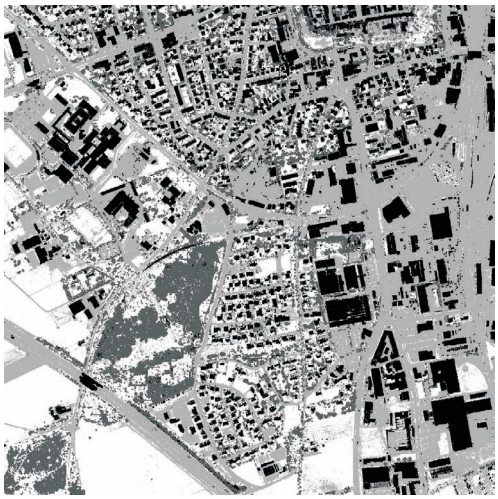

Figure 2. Initial Dempster-Shafer classification. White: grass-land, light grey: bare soil, dark grey: trees, black: buildings.

The results of Dempster-Shafer classification are presented in figure 2. Class "bare soil" mainly corresponds to streets, parking lots, railway tracks, sporting facilities, and fallow fields. Most of the trees are situated in the forests close to the centre and the south-western corner of the scene, along the streets, and in the backyards of the residential houses. The building borders are often classified as trees. There were problems separating buildings from trees in regions of shadow because in very dark areas the NDVI could not be determined with sufficient an accuracy. This was the reason why one large industrial building could not

be detected. Further problems occurred in areas of bridges. Morphologic filtering produces a DTM corresponding to the terrain below the bridge. The subsequent height differences ΔH indicate buildings or trees, whereas the surface roughness of the street and the NDVI indicate an area not covered by vegetation. The bare top of a small hill in the centre of the scene was erroneously classified as a building because it was cut off in a very early iteration of DTM generation. Both the surface roughness and the NDVI correctly suggested that the area was not covered by trees. There are also some single misclassified pixels of a class in the inside of larger areas of another class, because context was not considered in the classification process.

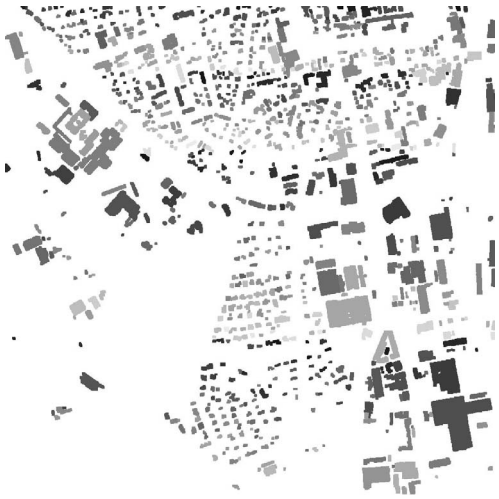

Figure 3. The final building label image.

After morphological opening and eliminating building candidate regions smaller than 30 m^2, the second Dempster-Shafer classification is carried out for 966 building candidate regions. Of these regions, 26 are classified as belonging to another class, so that altogether 940 building regions are detected. Figure 3 shows the final building label image after growing the building regions by morphologic closing. As no ground truth data are available for the test site, no comparison of these results to such data could be carried out. A coarse visual inspection revealed that only a few buildings (about 3%) were missed by the classification. These were small residential buildings with roofs consisting of very small planar faces or covering too small an area to be detected given the resolution of the LIDAR data. One large industrial building with many shadow areas on top of its roof was also missed. The percentage of false alarms seems to be in the same range as the rate of missed buildings, but no precise numbers can be given at this time. Problem areas are bridges, for reasons already described above, large overseas containers in the industrial areas, shadows, and small terrain structures not covered by vegetation.

5. 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 data fusion. The results achieved were satisfactory, although a more thorough evaluation including ground truth data to give precise detection rates still has to be done. In some cases, buildings and trees cannot be accurately separated, either because of shadows or because the resolution of the LIDAR data is not sufficient. Better results might be achievable with different definitions of the probability masses in the second classification step. We also want to use the results of our method to improve the quality of the DTM by eliminating points on the building roofs before applying robust linear prediction as described in [6].

Acknowledgements

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

References

[1] A. Brunn and U. Weidner. Extracting Buildings from Digital Surface Models. volume XXXII/3-4W2 of *IAPRS*, pages 27–34, 1997.

[2] W. Förstner. A Framework for Low Level Feature Extraction. volume II of *ECCV '94*, pages 383–394, 1994.

[3] N. Haala and C. Brenner. Extraction of buildings and trees in urban environments. *ISPRS Journal of Photogrammetry & Remote Sensing*, 54:130–137, 1999.

[4] L. Klein. *Sensor and data fusion, concepts and applications*. SPIE Optical Engineering Press, 1999.

[5] Y. H. Lu and J. Trinder. Data Fusion Applied to Automatic Building Extraction in 3D Reconstruction. Proc. ASPRS Conference, pages 114–122, Anchorage, Alaska, 2003.

[6] F. Rottensteiner and C. Briese. A New Method for Building Extraction in Urban Areas from High-Resolution LIDAR Data. volume XXXIV/3A of *IAPRSIS*, pages 295 – 301, 2002.

[7] F. Rottensteiner, J. Trinder, S. Clode, and K. Kubik. Building Detection Using LIDAR Data and Multispectral Images. Digital Image Computing - Techniques and Applications (DICTA) Vol. II, pages 673–682, Sydney, Australia, 2003.

[8] G. Shafer. *A mathematical theory of evidence*. Princeton University Press, 1976.

[9] U. Weidner. *Gebäudeerfassung aus digitalen Oberflächenmodellen*. PhD thesis, University of Bonn, 1997. DGK Volume 474.