

# A COMPARISON OF BAYESIAN AND EVIDENCE-BASED FUSION METHODS FOR AUTOMATED BUILDING DETECTION IN AERIAL DATA

K. Khoshelham<sup>a,\*</sup>, S. Nedkov<sup>a</sup>, C. Nardinocchi<sup>a,b</sup>

<sup>a</sup> Optical and Laser Remote Sensing, Delft University of Technology, 2629 HS Delft, The Netherlands  
- k.khoshelham@tudelft.nl

<sup>b</sup> DITS, University of Rome "La Sapienza", 00184 Roma, Italy - carla.nardinocchi@uniroma1.it

Commission VI, WG VI/4

**KEY WORDS:** Fusion, Building detection, Automation, Aerial image, Laser scanning

## ABSTRACT:

Automated approaches to building detection are of great importance in a number of different applications including map updating and monitoring of informal settlements. With the availability of multi-source aerial data in recent years, data fusion approaches to automated building detection have become more popular. In this paper, two data fusion methods, namely Bayesian and Dempster-Shafer, are evaluated for the detection of buildings in aerial image and laser range data, and their performances are compared. The results indicate that the Bayesian maximum likelihood method yields a higher detection rate, while the Dempster-Shafer method results in a lower false-positive rate. A comparison of the results in pixel level and object level reveals that both methods perform slightly better in object level.

## 1. INTRODUCTION

Automated approaches to building detection are of great importance in a number of applications, including map updating, city modeling and monitoring of informal settlements. Up-to-date maps are in high demand with the current widespread use of navigation systems. Map updating is a tedious task when it is performed manually by an operator. Each building has to be inspected in the map and in a recent aerial image (or a stereo pair). Changes are marked and the map is updated accordingly. This process is expensive and time consuming. Automation of this process will save time and cost, making it possible to do faster and more frequent map updating.

Using data from only one source of data often does not provide enough information to correctly detect buildings in an automated fashion. By fusing data from multiple sources the chances of correctly detecting buildings increase. Several methods of data fusion have been used for the detection of buildings from multi-source aerial data (Bartels and Wei,2006; Lu et al.,2006; Rottensteiner et al.,2004a; Walter,2004). While relatively successful application of these methods has been reported, a comparison of the performance of the methods is not available. The objective of this paper is to provide a comparison of the two main data fusion methods, namely Bayesian and Dempster-Shafer, as applied to the detection of buildings from multi-source aerial data.

The paper has the following outline: in the next section an overview of the previous research in the field of automated building detection is given. This includes building detection using one aerial image, stereo and multiple-overlap images, using height data and through the fusion of several sources. Section 3 provides a brief description of the Bayesian decision theory and the Dempster-Shafer evidence theory. In Section 4 feature extraction and evidence gathering for classification and

morphological post-processing for building detection are discussed. Experimental results and a comparison of the performance of the methods are presented in Section 5. Conclusions are drawn in Section 5.

## 2. RELATED WORK

Building detection from aerial images has been a hot topic since the early 1990's. Early approaches were based on a single image. Buildings were detected by making use of their shadows (Lin and Nevatia,1996). Shadows have the disadvantage of not being visible in all situations. In addition, they can obscure one another thereby forming complex shapes, which do not resemble the original buildings. Lin and Nevatia,(1996) furthermore assume that the roofs of the buildings are rectilinear, flat, and that the shadows cast by the buildings fall on flat ground. This is clearly not the case in many urban areas. Using more than one image supplies more information and other methods of detection can be applied. Fischer et al.,(1998) use multiple images to recognize building features (points, lines and regions). These features are then used to construct building corners, wings, and faces, which in turn are combined to reconstruct a building. Fradkin et al.,(2001) use multiple images to detect facades of buildings. Several images from different angles are taken so that the problem of occlusion is reduced to some extent. After the facade has been found the rest of the building is detected. Contrary to roof detection (Khoshelham et al.,2005; Muller and Zaum,2005), facades have less area from which they can be detected, making their detection troublesome.

Using several images it is possible to construct a height model of the scene. Weidner and Forstner,(1995) and Brunn and Weidner,(1997) use the height differences between buildings and the ground to make a guess about what is a building and what is not. Using this method, trees can be classified as

---

\* Corresponding author.

buildings if they have the same height as buildings. A height model of the scene can also be generated using an airborne laser scanner. Vosselman,(1999) uses laser scanner data to generate a Delaunay triangulation to reconstruct the rooftops and thus the buildings.

With the availability of multi-source aerial data in recent years, data fusion approaches to building detection have increasingly attracted attention. One of the common approaches to data fusion for classification of objects is based on the Bayesian decision theory. Walter,(2004) applies Bayesian maximum likelihood method to object-based classification of multi-spectral data. Rottensteiner et al.,(2004b) fuse height data with NDVI from multispectral images based on Dempster-Shafer theory. Lu et al.,(2006) use a similar method to detect buildings in multi-source aerial data based on the Dempster-Shafer theory.

### 3. BAYESIAN AND EVIDENCE-BASED DATA FUSION FOR BUILDING DETECTION

#### 3.1 Bayesian approach

In a typical data fusion strategy, feature vectors are extracted from the multi-source data, and a decision is made for each feature vector as to what class of object it belongs to. In the Bayesian fusion, a decision is made on the basis of maximizing the likelihood of a feature vector  $x$  being belonged to a class  $w_j$ . This is realized by evaluating a decision function for each feature and each class (Duda et al., 2001):

$$d_j(x) = p(x/w_j) \cdot P(w_j) \quad (1)$$

where the conditional probability  $p(x/w_j)$  is the probability of a feature vector  $x$  when drawn from the class  $w_j$ , and  $P(w_j)$  is the prior probability of the class  $w_j$ . The pixel or object with the feature vector  $x$  is then assigned to the class  $w_j$  if:

$$d_j(x) > d_i(x) \quad \forall i \neq j \quad (2)$$

The probabilities in Eq. (1) are derived from the training data. Feature vectors are often assumed to have a Gaussian distribution; thus,  $p(x/w_j)$  is replaced with a multi-dimensional Gaussian function with parameters  $\mu$  and  $\Sigma$  (mean and covariance respectively). The decision function can then be expressed as:

$$d_j(x) = -\frac{1}{2}(x - \mu_j)^T \Sigma_j^{-1} (x - \mu_j) - \frac{1}{2} \log |\Sigma_j| + \log P(w_j) \quad (3)$$

Classification of features based on the decision function given in Eq. (3) is referred to as the maximum likelihood method. A simple case of the maximum likelihood method is when an assumption can be made that the features in all classes are independent and have the same variance. Further, if it can be assumed that the prior probabilities of all classes are the same, the decision function in Eq. (3) will simplify to:

$$d_j(x) = -\frac{1}{2}(x - \mu_j)^T (x - \mu_j) \quad (4)$$

which suggests that instead of maximizing the decision function, a classification of feature vectors can be performed on the basis of minimizing the distance of each feature to the means of the

classes. Accordingly, the method is referred to as the minimum distance classification.

#### 3.2 Dempster-Shafer approach

Dempster-Shafer data fusion approach is based on making decisions according to available evidence for each object class. Each feature is seen as a piece of evidence that provides a certain degree of belief in each class hypothesis (Gordon and Shortliffe, 1990). Hypotheses include not only all classes but also any union of the classes. The effect of a piece of evidence to the hypotheses is represented by a probability mass assignment function  $m$ . The amount of belief to a hypothesis  $A$  is represented by a belief function:

$$Bel(A) = \sum_{B \subseteq A} m(B) \quad (5)$$

which is the sum of the mass probabilities assigned to all subsets of  $A$  by  $m$ . When two or more evidences are available, the probability masses assigned to the hypotheses are combined using the following combination rule:

$$m(A) = \frac{\sum_{\substack{i,j \\ B_i \cap C_j = A}} m_1(B_i) \cdot m_2(C_j)}{1 - k} \quad (6)$$

where  $1-k$  is a normalization factor in which  $k$  is the sum of all non-zero values assigned to the null set hypothesis  $\emptyset$ . The decision on the class of a feature can be made based on a maximum belief decision rule, which assigns a feature to a class  $A$  if the total amount of belief supporting  $A$  is larger than that supporting its negation:

$$Bel(A) \geq Bel(\bar{A}) \quad (7)$$

## 4. EXPERIMENTS AND RESULTS

#### 4.1 Experimental setup

Experiments were conducted to evaluate the performance of the Bayesian and evidence-based fusion in building detection. The study area was a suburban part of the city of Memmingen, south of Germany.

**4.1.1 Data:** The available data for the experiments included airborne laser range data containing first pulse and last pulse DSMs with a density of 1 point per  $m^2$ , and orthorectified aerial imagery in visible and near infrared channels with a ground resolution of 0.5m. In addition, a DTM of the scene was available in which buildings and other objects were filtered out. Fig. 1 depicts the color infrared orthoimage and the first pulse laser range image of the study area.

**4.1.2 Classes:** For the classification of objects in the study area four classes were considered: building (B), tree (T), bare land (L) and grass (G). The main object of interest was building, and in the final evaluation only buildings were taken into account.

**4.1.3 Features:** Three types of features were extracted from the data:

- height difference between the first pulse laser DSM and the last pulse laser DSM (DSM<sub>fe</sub>-DSM<sub>le</sub>);
- height difference between the last pulse laser DSM and the DTM (DSM<sub>le</sub>-DTM);
- Normalized Difference Vegetation Index (NDVI) derived from the red and near infrared image channels.

In the Bayesian approach, the extracted features were directly used for evaluating the decision function. To compute the mean and covariance matrix for each class a set of training regions was specified in the data. Fig. 2.A shows the training regions.

In the Dempster-Shafer approach, features were used to obtain evidence (probability mass) for each class hypothesis. For the derivation of evidence we developed linear evidence assignment functions based on the training data. These functions were tuned by trial and error to yield best separation of the objects. Fig. 3 illustrates the evidence assignment functions.

The combination of evidences derived from different features was carried out using the combination rule given in Eq. (6). Table 1 and Table 2 demonstrate the evidence combination process where  $s$  is the evidence derived from the height difference between the last pulse DSM and the DTM,  $t$  is the evidence derived from the height difference between the first and last pulse DSMs, and  $u$  is the evidence derived from the NDVI. Table 3 summarizes the final evidence values computed for the four classes.

Both the Bayesian and Dempster-Shafer methods were applied in pixel level and in object level. In pixel level, features and evidence values were computed for each individual pixel. In object level, a segmentation of the color infrared orthoimage was first obtained, and the mean of features within each segment was used in the computations. The segmented orthoimage is depicted in Fig. 2.B. As can be seen, the image is slightly oversegmented so that overgrown regions are avoided.

**4.1.4 Detection of buildings:** As mentioned before, buildings are the main object of interest in this research. Therefore, to evaluate the performance of the fusion methods in the context of building detection a binary building image was obtained from the classification results. A morphological opening operation was applied to clean this binary building image from small objects that were identified as building. This process was followed by a morphological reconstruction operation to retrieve the building boundaries that were smoothed out in the opening process. Fig. 4 illustrates the effect of this post-processing in a sample binary building image.

**4.2 Results**

The Bayesian and Dempster-Shafer methods were applied to the data in both pixel level and object level. In the application of Bayesian maximum likelihood method, we noticed that, the variance of the height difference between the first and last pulse laser data in the building training regions was considerably small. To examine the influence of small variance, and consequently nearly singular covariance matrix, we also classified the data with the minimum distance method. Fig. 5 shows the classification results as well as the detected buildings

obtained by applying the three methods to the data in pixel level. Fig. 6 shows the object-level results. As can be seen, the three methods yield slightly different classifications of the data in the four predefined object classes methods; however unlike the Bayesian maximum likelihood and minimum distance methods, the classification results of the Dempster-Shafer method include also pixels and objects that are not assigned to any of the four classes (shown in black in the lower left images of Fig. 5 and Fig. 6). For the evaluation, these unclassified pixels were considered not-building.

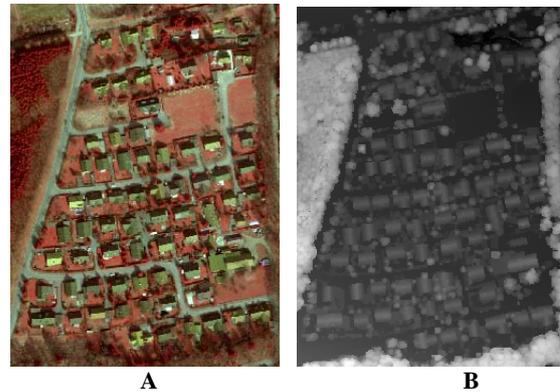


Fig. 1. A. Color infrared aerial orthoimage of the study area; B First pulse laser range image of the area.

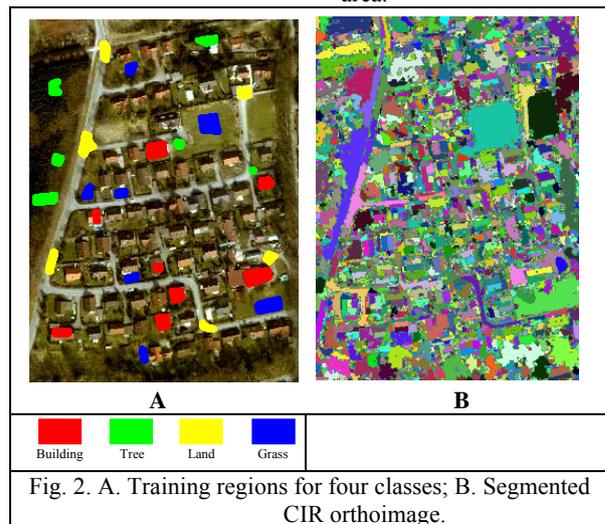


Fig. 2. A. Training regions for four classes; B. Segmented CIR orthoimage.

The evaluation of the performance of the methods was carried out based on a set of ground truth data that contained building boundaries extracted manually from the RGB orthoimage. Fig. 7 depicts these reference building boundaries. Three performance measures were obtained by comparing the buildings detected using each method and the reference data:

- Detection rate: the ratio of the number of pixels correctly detected as building to the total number of building pixels according to the reference data;
- False positive: the ratio of the number of pixels wrongly detected as building to the total number of not-building pixels according to the reference data;
- False negative: the ratio of the number of pixels wrongly detected as not-building (missed building pixels) to the total number of building pixels according to the reference data.

Table 4 summarizes the performance measures obtained by applying each of the three methods to the data in pixel level. The performance measures of the methods in object level are given in Table 5. As can be seen, the Bayesian maximum likelihood method yields the highest detection rate and the lowest false negative rate in both pixel level and object level. The Dempster-Shafer method and the minimum distance method perform better in terms of the false positive rate. For both methods the false positive rate is about 5 times lower than that of the maximum likelihood method. A comparison of the performance of the methods in pixel level and in object level reveals that the detection rates and the false negative rates are slightly improved for all methods in object level, whereas the false positive rates are still better in pixel level.

	s	{B} (s)	{T,L,G} (1-s)
t			
{T} (t)		{∅} st	{T} t(1-s)
{B,L,G} (1-t)		{B} s(1-t)	{L,G} (1-s)(1-t)

Table 1. Combination of evidence s (from DSMle-DTM) with t (from DSMfe-DSMle).

sot	{B}	{T}	{L,G}
u	$\frac{s(1-t)}{(1-st)}$	$\frac{t(1-s)}{(1-st)}$	$\frac{(1-s)(1-t)}{(1-st)}$
{T,G}	{∅}	{T}	{G}
(u)	$\frac{su(1-t)}{(1-st)}$	$\frac{tu(1-s)}{(1-st)}$	$\frac{u(1-s)(1-t)}{(1-st)}$
{B,L}	{B}	{∅}	{L}
(1-u)	$\frac{s(1-t)(1-u)}{(1-st)}$	$\frac{t(1-s)(1-u)}{(1-st)}$	$\frac{(1-s)(1-t)(1-u)t}{(1-st)}$

Table 2. Combination of evidence u (from NDVI) with sot (combination of s and t).

Class hypothesis	Combined evidence (sotou)
{B}	$\frac{s(1-t)(1-u)}{(1-t+tu-su)}$
{T}	$\frac{(1-s)tu}{(1-t+tu-su)}$
{L}	$\frac{(1-s)(1-t)(1-u)}{(1-t+tu-su)}$
{G}	$\frac{(1-s)(1-t)u}{(1-t+tu-su)}$

Table 3. Combined evidences for simple class hypotheses.

By visual inspection of the results in Fig. 5 and Fig. 6, it can be observed that building objects detected by the application of the methods in object level are relatively larger than those detected in pixel level. A superimposition of the detected buildings on the reference boundary map showed that in pixel-based results many building pixels at the boundaries of buildings were missed. This explains the better detection rate and false negative rate of the object-based results.

An examination of the detection results also suggests that when the detected buildings are to be compared against a map for the purpose of change detection, the Dempster-Shafer results provide better signals for an operator, as compared to the Bayesian results, due to the much lower rate of false positive signals.

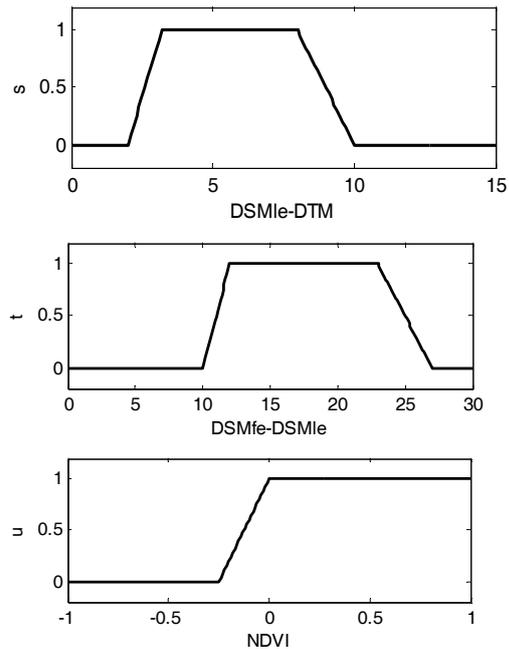


Fig. 3. Evidence assignment functions for the Dempster-Shafer method.

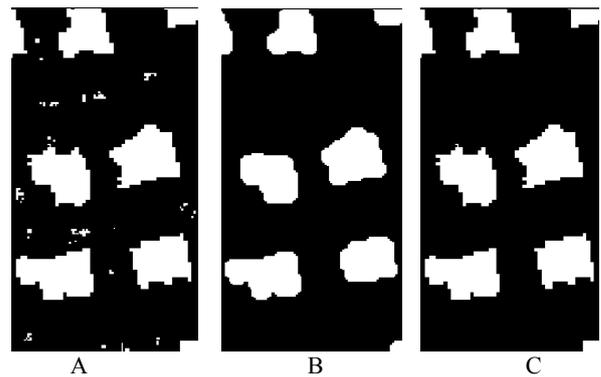


Fig. 4. Cleaning of the binary building image using morphological operations. A. Binary building image; B. Morphological opening removes small objects, but also smoothes out building boundaries; C. Morphological reconstruction retrieves the building boundaries.

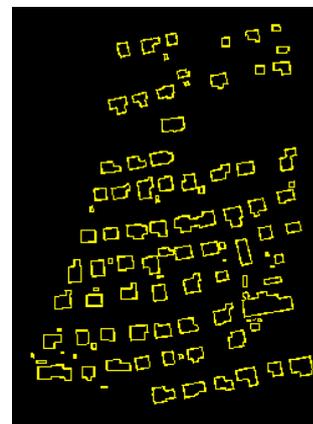
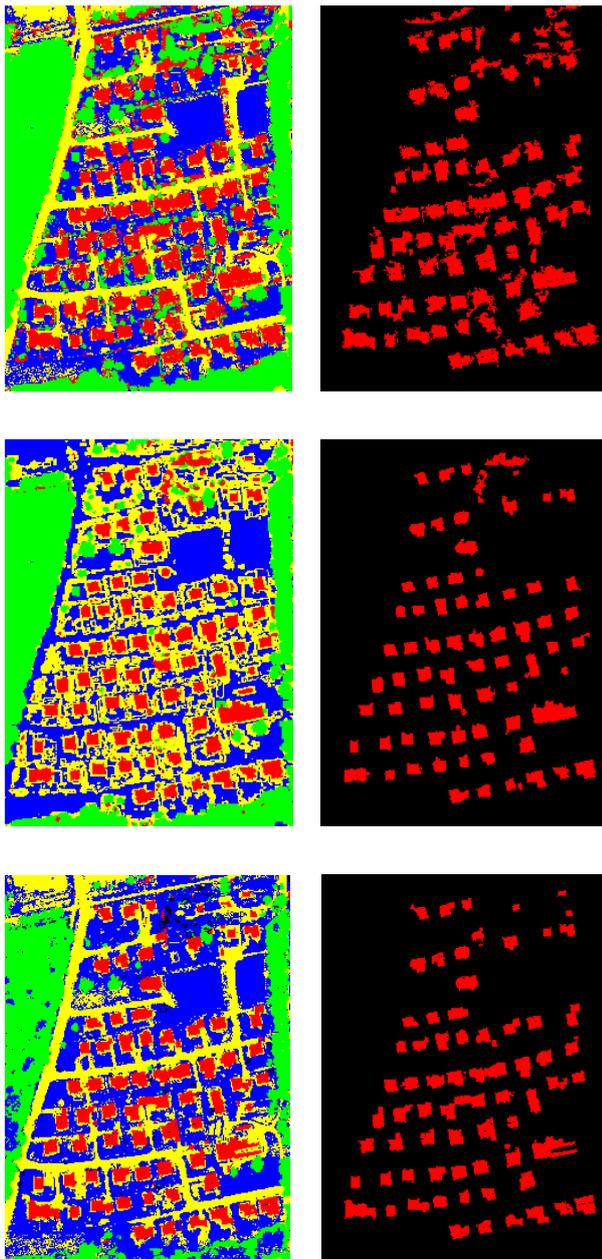


Fig. 7. Manually extracted building boundaries used as reference data for the evaluation of detected buildings.

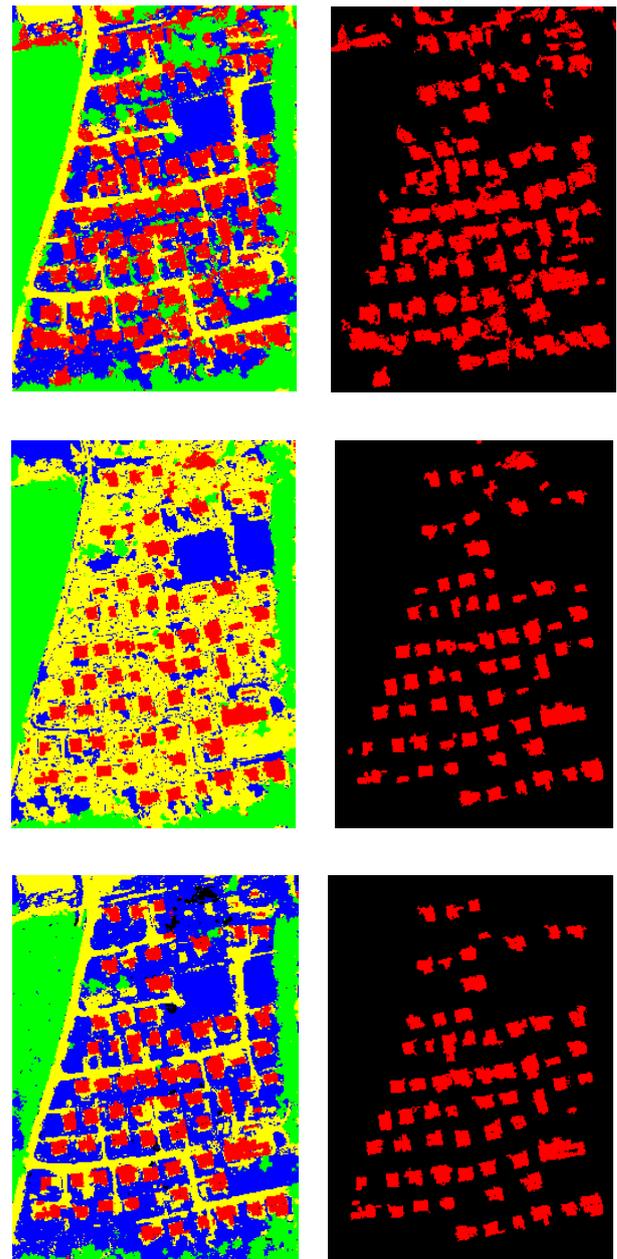


Building Tree Land Grass

Fig. 5. Classification results (left column) and detected buildings (right column) obtained by applying the algorithms in pixel level; Top row: results of maximum likelihood method; Middle row: results of minimum distance method; Bottom row: results of Dempster-Shafer method.

	Detection rate	False Positive	False Negative
ML_pix	0.87	0.05	0.13
MD_pix	0.68	0.01	0.32
DS_pix	0.75	0.01	0.25

Table 4. Performance measures for detection in pixel level using the three methods.



Building Tree Land Grass

Fig. 6. Classification results (left column) and detected buildings (right column) obtained by applying the algorithms in object level; Top row: results of maximum likelihood method; Middle row: results of minimum distance method; Bottom row: results of Dempster-Shafer method.

	Detection rate	False Positive	False Negative
ML_obj	0.95	0.11	0.05
MD_obj	0.71	0.02	0.29
DS_obj	0.78	0.02	0.22

Table 5. Performance measures for detection in object level using the three methods.

## 5. CONCLUSIONS

Automated approaches to building detection are important in the updating of cadastral maps and monitoring of informal settlements. In this paper, a comparative analysis of two data fusion and classification approaches, namely Bayesian and Dempster-Shafer, as applied to automated building detection in aerial data was presented. Results showed that both methods perform slightly better in object level than in pixel level. A comparison of the performance of the methods revealed that the Bayesian maximum likelihood method yields a higher detection rate, as compared to the minimum distance method and the Dempster-Shafer method; however, the rate of pixels wrongly detected as building is also higher in the Bayesian method. In practice, the crucial measure in the evaluation a detection method is the rate of missed building pixels. In this respect, the Bayesian maximum likelihood method was found to have a better performance; however, the missed pixels in the Dempster-Shafer method were found to be mostly at the boundaries of buildings. Therefore, the higher rate of the missed building pixels in the Dempster-Shafer method should not be seen as a critical drawback of the method.

In this research, we trained the Dempster-Shafer evidence assignment functions using the training regions, and on the basis of a trial and error procedure. Further research can be focused on developing more elaborated training algorithms for the assignment of evidence in the Dempster-Shafer method.

## ACKNOWLEDGEMENTS

The authors would like to thank the Toposys company for providing the dataset that was used in the experiments.

## REFERENCES

- Bartels, M. and Wei, H.,2006. Maximum likelihood classification of LIDAR data incorporating multiple co-registered bands, 4th International Workshop on Pattern Recognition in Remote Sensing in conjunction with the 18th International Conference on Pattern Recognition, Hong Kong, pp. 17-20.
- Brunn, A. and Weidner, U.,1997. Extracting buildings from digital surface models, ISPRS Workshop on 3D Reconstruction and Modelling of Topographic Objects, Stuttgart, pp. 27-34.
- Duda, R.O., Hart, P.E. and Stork, D.G.,2001. Pattern classification, second edition. Wiley, New York, 654 pp.
- Fischer, A. et al.,1998. Extracting buildings from aerial images using hierarchical aggregation in 2D and 3D. *Computer Vision and Image Understanding*, 72(2): 185-203.
- Fradkin, M., Maitre, H. and Roux, M., 2001. Building detection from multiple aerial images in dense urban areas. *Computer Vision and Image Understanding*, 82: 181-207.
- Gordon, J. and Shortliffe, E.H.,1990. The Dempster-Shafer theory of evidence. In: G. Shafer and J. Pearl (Editors), *Readings in uncertain reasoning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 768.
- Haala, N. and Brenner, C.,1998. Interpretation of urban surface models using 2D building information. *Computer Vision and Image Understanding*, 72(2): 204-214.
- Huertas, A., Lin, C. and Nevatia, R.,1993. Detection of buildings from monocular views of aerial scenes using perceptual organization and shadows, ARPA image understanding workshop, Washington, DC, pp. 253-260.
- Khoshelham, K., Li, Z.L. and King, B.,2005. A split-and-merge technique for automated reconstruction of roof planes. *Photogrammetric Engineering and Remote Sensing*, 71(7): 855-862.
- Lin, C. and Nevatia, R.,1996. Buildings detection and description from monocular aerial images, ARPA Image Understanding Workshop, Palm Springs, CA.
- Lu, Y.H., Trinder, J.C. and Kubik, K.,2006. Automatic building detection using the Dempster-Shafer algorithm. *Photogrammetric Engineering and Remote Sensing*, 72(4): 395-403.
- Muller, S. and Zaum, D.,2005. Robust building detection in aerial images, CMRT '05, Vienna, pp. 143-148.
- Nevatia, R., Lin, C. and Huertas, A.,1997. A system for building detection from aerial images. In: A. Gruen, E. Baltsavias and O. Henricsson (Editors), *Automatic extraction of man-made objects from aerial images (II)*. Birkhauser Verlag, Basel, pp. 77-86.
- Rottensteiner, F., Trinder, J., Clode, S. and Kubik, K.,2004a. Fusing airborne laser scanner data and aerial imagery for the automatic extraction of buildings in densely built-up areas, *International Archives of Photogrammetry and Remote Sensing*, Vol. XXXV, Part B3, Istanbul, pp. 512-517.
- Rottensteiner, F., Trinder, J., Clode, S., Kubik, K. and Lovell, B.,2004b. Building detection by Dempster-Shafer fusion of LIDAR data and multispectral aerial imagery, *Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)*, Cambridge, United Kingdom.
- Vosselman, G.,1999. Building reconstruction using planar faces in very high density height data, *ISPRS Conference on Automatic Extraction of GIS Objects from Digital Imagery*, Munich, pp. 87-92.
- Walter, V.,2004. Object-based classification of remote sensing data for change detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(2004): 225-238.
- Weidner, U. and Forstner, W.,1995. Towards automatic building extraction from high resolution digital elevation models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 50(4): 38-49.