A LEARNING DEMPSTER-SHAFER MODEL FOR AUTOMATED BUILDING DETECTION

Kourosh Khoshelham
Optical and Laser Remote Sensing, Delft University of Technology, The Netherlands
Email: k.khoshelham@tudelft.nl

Carla Nardinocchi
DITS, University of Rome "La Sapienza", 00184 Roma, Italy
Email: carla.nardinocchi@uniroma1.it

Abstract: This paper presents a learning Dempster-Shafer model for the detection of buildings in aerial image and range data. The process of evidence assignment in the Dempster-Shafer method is implemented through membership functions in an adaptive-network-based fuzzy inference system, where a back propagation learning rule is employed to tune the evidence assignment functions using training samples. The advantage of this method is that it incorporates our knowledge about various features that can be extracted from multi-source aerial data, and the evidence that these features provide for buildings and other objects in urban and suburban areas. Experimental results show that the proposed learning model improves the performance of the Dempster-Shafer classifier in detecting buildings in multi-source aerial data.

1. Introduction

Automated building detection in aerial data is a complex problem that has challenged researchers in the field of photogrammetry and computer vision for many years. In recent years, a growing interest has been shown in the application of the Dempster-Shafer theory of evidence (Dempster, 1967; Shafer, 1976) to automated building detection in multi-source aerial data (Frontoni et al., 2008; Khoshelham et al., 2008; Lu et al., 2006; Rottensteiner et al., 2005; Rottensteiner et al., 2004). The Dempster-Shafer approach to building detection is based on fusing pieces of uncertain evidence derived from various sources of data. Evidences are basic probabilities that support the belief that an object belongs to a certain class hypothesis. The treatment of uncertainty in data is a main advantage of the Dempster-Shafer approach over previous probabilistic methods (Gordon and Shortliffe, 1990).

A major problem associated with the Dempster-Shafer method is the assignment of basic probabilities to features extracted from the data. Often, the probabilities are assigned by means of some mass probability assignment functions, which are defined based on an expert's knowledge about the object. Previous attempts to devise a learning mechanism wherein the mass probability assignment functions can be automatically tuned and learnt from training data mainly focus on combining the Dempster-Shafer model with non-deterministic methods such as fuzzy reasoning (Binaghi and Madella, 1999), neural networks (Basir et al., 2005; Binaghi et al., 2000), and genetic algorithms (Sosnowski and Walijewski, 2005). These combinations lead to general data-driven classifiers which do not necessarily incorporate available knowledge of the specific problem at hand.

In this paper, we propose a learning Dempster-Shafer model for the detection of buildings in aerial range and image data. In the proposed model, the mass probability assignment functions are treated as membership functions in an adaptive-network-based fuzzy inference system (Jang, 1993), where a back propagation learning rule is employed to adapt the mass probability assignment functions with respect to given training samples. The advantage of the proposed method is that it incorporates our knowledge about various features that can be extracted from multi-source aerial data, and the evidence that these features provide for buildings and other objects in urban and suburban areas. We show that the proposed learning model improves the performance of the Dempster-Shafer classifier in detecting buildings in multi-source aerial data.

The paper proceeds with a description of the Dempster-Shafer approach to building detection in Section 2. In Section 3, the learning model for the tuning of the mass probability assignment functions using training samples is described. Experimental results and comparisons with other methods are presented in Section 4. Conclusions appear in Section 5.

2. Building Detection Based on Evidence

The Dempster-Shafer approach to building detection relies on gathering evidence from the data for the hypothesized classes. The evidence that a feature extracted from the data provides for a class hypothesis is represented by mass probabilities in the range [0,1] assigned to class hypotheses such that the mass probabilities sum to one. If a feature provides evidence against a class hypothesis, a large mass probability (close to 1) may be given to the negation of that hypothesis, and a small probability (close to 0) to the union of all hypotheses. The mass probabilities derived from multiple features are combined using Dempster's rule of combination:

$$\mathbf{m}(B) = \frac{\sum_{C_i \cap C_j = A} m_1(C_i) \cdot m_2(C_j)}{1 - \sum_{C_i \cap C_j = \varnothing} m_1(C_i) \cdot m_2(C_j)} \tag{1}$$

where $m_1(C_i)$ and $m_2(C_j)$ are the mass probabilities (from two features) assigned respectively to the class hypotheses *Ci* and *Cj*, and $\mathbf{m}(B)$ is the combined evidence for the class hypothesis *B*, which is the intersection of the two sets *Ci* and *Cj*. The sum of the combined evidences assigned to all subsets of a class hypothesis gives rise to an amount of belief for that hypothesis:

$$\mathbf{Bel}(A) = \sum_{B \subseteq A} \mathbf{m}(B) \tag{2}$$

The computed belief is the basis for classification. A pixel or object is assigned to class *A* if the amount of belief in support of *A* is larger than that supporting its negation and the other single class hypotheses.

Based on the Dempster-Shafer theory we develop a model for the classification of multi-source aerial data into four classes: Building (B), Tree (T), Land (L), and Grass (G). The aerial data should include imagery in red and near-infrared bands as well as laser range data in the form of a first echo digital surface model (DSM), a last echo DSM, and a filtered version of the last echo DSM in which all the objects on the surface of the terrain are flattened out.

This latter dataset is a representation of the terrain only and is referred to as DTM. The Dempster-Shafer classification model works with three features extracted from the data:

- *le-dtm*: height difference between the last echo DSM and the DTM;
- *fe-le*: height difference between the first echo DSM and the last echo DSM;
- *ndvi*: normalized difference vegetation index derived from the red and near infrared image bands.

The basic assumption in the proposed Dempster-Shafer classification model is that *le-dtm* provides evidence for the class building (corresponding to probability s), while *fe-le* provides evidence for the class tree (corresponding to probability t), and *ndvi* is evidence for both tree and grass (corresponding to probability u). These assumptions are represented in three initial mass probability assignment functions that determine roughly the amount of mass probabilities that each feature provides for the class hypotheses. The initial mass probability assignment functions can be easily constructed based on available knowledge about the study area, such as the minimum and maximum height of the buildings and trees. Figure 1 shows an example of initial mass probability assignment functions that we have used in this research.

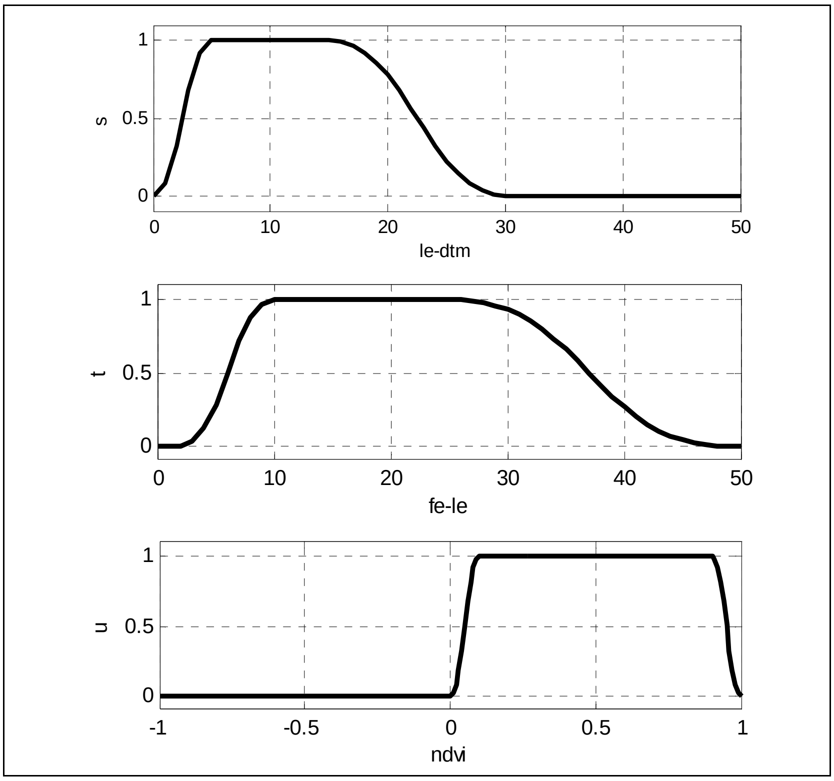

Figure 1: The initial mass probability assignment functions.

Having considered four classes of objects present in the data, a set of 16 class hypotheses is obtained. The mass probabilities derived from each feature are assigned to the class hypotheses using the probability assignment functions shown in Figure 1. The mass probabilities are then combined using the combination rule (Eq. 1) and an amount of belief is computed for each class hypothesis. Table 1 summarizes the class hypotheses, and the assignment and combination of the mass probabilities.

Class hypothesis	Mass probabilities			Combination (s, t)	Combination (s, t, u)
	le-dtm	fe-le	ndvi		
{Ø}	0	0	0	0	0
{B}	**s**	0	0	**s(1-t)/(1-st)**	**s(1-t)(1-3u)/(1-t-u-2su+3tu)**
{T}	0	**t**	**u**	**(1-s)t/(1-st)**	**2(1-s)tu/(1-t-u-2su+3tu)**
{L}	0	0	0	0	**(1-s)(1-t)(1-3u)/(1-t-u-2su+3tu)**
{G}	0	0	**u**	0	**2(1-s)(1-t)u/(1-t-u-2su+3tu)**
{B, T}	0	0	0	0	0
{B, L}	0	0	**1-3u**	0	0
{B, G}	0	0	0	0	0
{T, L}	0	0	0	0	0
{T, G}	0	0	**u**	0	0
{L, G}	0	0	0	**(1-s)(1-t)/(1-st)**	0
{B, T, L}	0	0	0	0	0
{B, T, G}	0	0	0	0	0
{B, L, G}	0	**1-t**	0	0	0
{T, L, G}	**1-s**	0	0	0	0
{B, T, L, G}	0	0	0	0	0

Table 1: The assignment and combination of the mass probabilities.

3. Neuro-Fuzzy Learning Model

The tuning of the mass probability assignment functions is carried out in two steps: in the first step, the initial mass probability assignment functions are imported as membership functions in a fuzzy inference system; in the second step, the membership functions are tuned in an adaptive feedforward neural network using a set of training samples. In the following, these processes are described in more detail.

3.1. Fuzzy Reasoning for Evidence Assignment

The process of mass probability assignment is in essence a fuzzy reasoning process. We, therefore, implement this process in a fuzzy inference system (Zadeh, 1973), wherein the initial mass probability assignment functions are imported as membership functions. The core of this fuzzy inference system is a set of four linguistic fuzzy rules, which translate our knowledge of the features and the characteristics of the objects into a mathematical representation. The rules establish the relation between the features as input and the object classes as output. Figure 2 (left) demonstrates the fuzzy rules for evidence assignment. The conditional statement that a feature provides evidence for an object class is fuzzified by means of a membership function. The membership functions specify the amount of evidence derived from the features in the exact same way as the mass probability assignment functions do in the Dempster-Shafer approach. Therefore, the initial probability assignment functions are used in the fuzzy inference system as the input membership functions. The output of each rule is evaluated using a Sugeno-type membership function (Sugeno, 1985), which associates each object class with a constant number in the range [0 1] (Here 0.1 for B, 0.2 for T, 0.3 for L, and 0.4 for G). Figure 2 (right) depicts the membership functions corresponding to the inputs and outputs of the four fuzzy rules. The fuzzy inference system maps the input features to an output object class by evaluating each rule and aggregating the outputs.

1. If *le-dtm* is evidence for B, then object is B;
2. If *fe-le* is evidence for T, then object is T;
3. If *le-dtm* is not evidence for B, and *fe-le* is not evidence for T, and *ndvi* is evidence for G&T, then object is G;
4. If *le-dtm* is not evidence for B, and *fe-le* is not evidence for T, and *ndvi* is not evidence for G&T, then object is L.

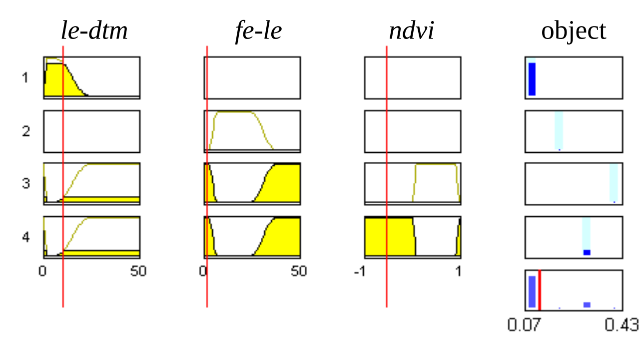

Figure 2: Fuzzy inference system for evidence assignment; left: the fuzzy rules; right: the corresponding membership functions.

3.2. Learning within an Adaptive Network

To Tune the membership functions, we adopt the back propagation learning rule using the gradient descent method as described by Jang (1993), which is based on constructing an adaptive neural network corresponding to a fuzzy inference system. An adaptive neural network is a set of (adaptive) nodes and connections, which map an input set of variables to an output, similar to a fuzzy inference system. The input and output variables, membership functions, and operations within the fuzzy inference system are represented by nodes and connections in the network architecture. Figure 3 shows the adaptive network construction corresponding to our fuzzy inference system for evidence assignment.

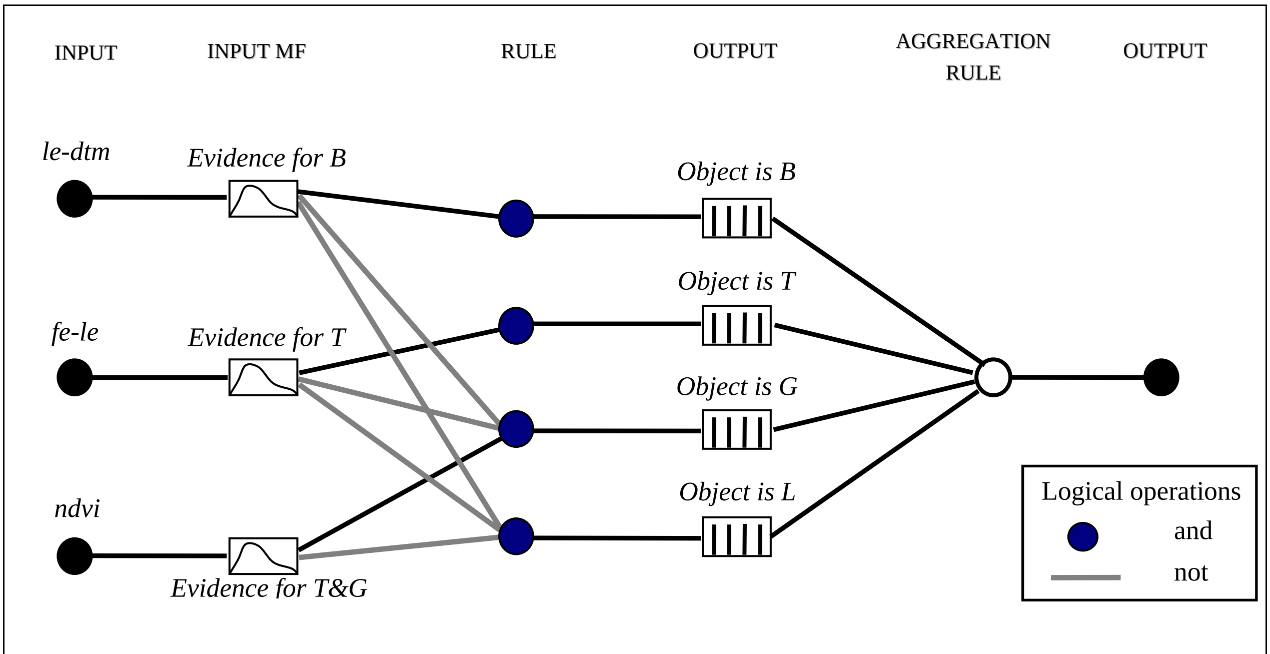

Figure 3: Adaptive network for the learning of the probability assignment functions.

The membership functions are contained within the adaptive nodes in the first layer of the network. An adaptive node is one whose output is a parametric function of the inputs to the node. The goal of the learning algorithm is to adjust the parameters of the adaptive node functions using a set of training samples such that an error measure is minimized. Each training sample is a set of input-output values, where the inputs are features extracted from the data, and the output value represents an object class. The error measure is usually the sum of the squared differences between the output values from the training samples and the outputs of the network (given inputs from the training set).

The gradient descent algorithm for adjusting the parameters of the membership functions is an iterative learning procedure. Each epoch of the iteration consists of a forward pass and a backward pass. In the forward pass, the input values from the training set are presented to the network, and a vector of outputs is computed. A total error measure is obtained as the sum of the squared differences between the entries of this output vector and the actual output values given in the training set. In the backward pass, this total error is back propagated to partial error rates at the output of each internal node. From the partial error rates at the output of the adaptive nodes in the first layer of the network an update formula is obtained for the adjustment of the parameters of the membership functions. With the updated parameters a new error measure is obtained, which is again back propagated to the internal adaptive nodes to update again the parameters of the membership functions. The error back propagation and update procedures are iterated until a minimum acceptable total error is obtained, or a certain number of iterations is reached. More details about the back propagation rule and the update formula can be found in (Jang, 1993).

4. Experiments

To experiment with the learning Dempster-Shafer model a set of aerial image and laser data acquired over a built-up area at the centre of Mannheim, Germany was used. The laser data included first echo and last echo recordings and a filtered version of the last echo data representing the terrain height. All laser data were interpolated at a regular grid with 0.5m point spacing. The image data contained orthorectified images in visible and near infrared spectral bands with a ground pixel size of 0.25m. To assess the performance of the method in detecting buildings a reference map of the area containing all building boundaries was manually compiled from the data. Figure 4 shows the aerial image, the last echo DSM, and the manually measured reference map of the area.

Figure 4: The dataset used in the experiment; Left: aerial orthoimage; Middle: last echo DSM; Right: reference map of the area.

Three features were extracted from the data as described in Section 2. These features were input to the Dempster-Shafer model with initial probability assignment functions shown in Figure 1. For the tuning of these probability assignment functions a set of training samples was obtained from the data. Figure 4 (left) depicts the regions, where training samples were collected, superimposed on the aerial orthoimage. The learning of the initial probability assignment functions was carried out by feeding the training samples to the adaptive network shown in Figure 3, and performing the iterative gradient descent procedure. Figure 5 illustrates the result of the learning process. As can be seen, within the first 150 iterations the total error reduces almost linearly to about half of its initial value. After 150 iterations the convergence rate slows down and the computed updates become very small, implying that the iterations can be terminated.

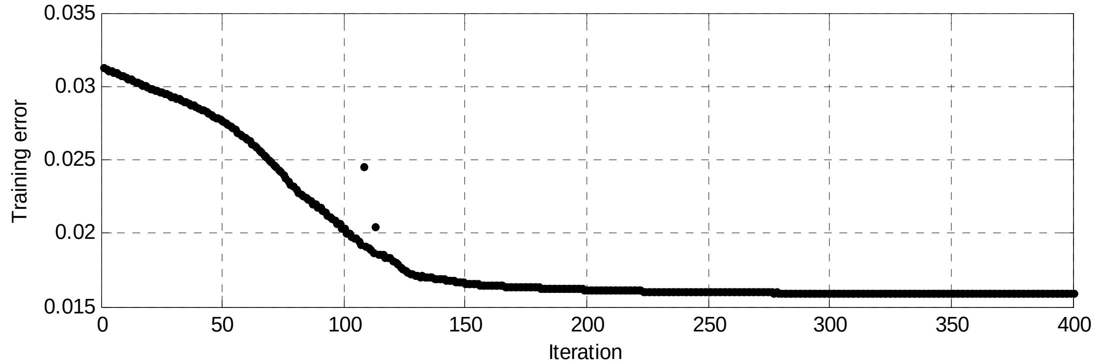

Figure 5: Reduction of the total error rate in the iterative gradient descent learning process.

The reduced error of the learned adaptive network is evident also from the performance of the network in the classification of the input samples. Figure 6 demonstrates the performance of the learned adaptive network, given a set of 100 check samples which were not used for the learning. As can be seen, the adaptive network yields a correct classification for 98% of the samples.

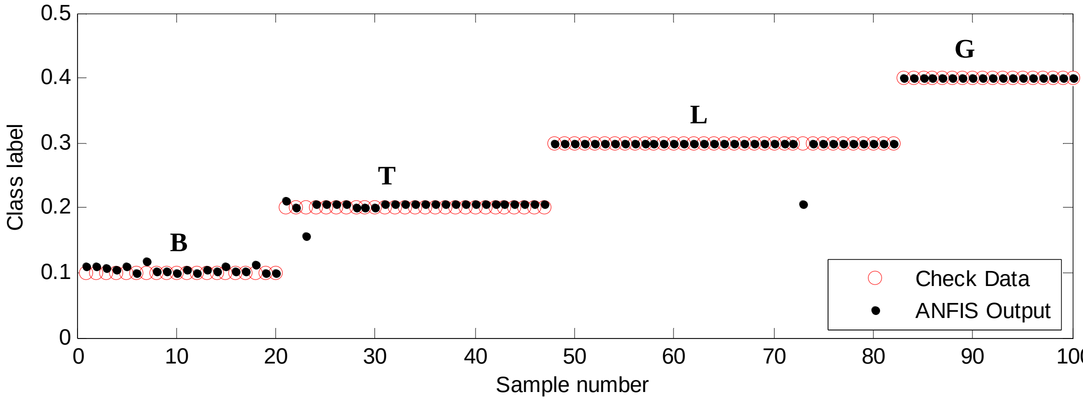

Figure 6: Performance of the learned adaptive network in the classification of check samples.

A complete classification of the dataset shown in Figure 4 was performed by using the updated membership functions of the learned adaptive network as probability assignment functions in the Dempster-Shafer classifier. Buildings were extracted from the classification result into a binary building image, and an evaluation of this building image was carried out using the reference map and based on the following error measures:

- *False Negative Rate (FNR)*: The ratio of the number of building pixels wrongly classified as other objects to the total number of building pixels;

- *False Positive Rate (FPR)*: The ratio of the number of non-building pixels wrongly classified as building to the total number of non-building pixels;
- *Unclassified Positive Rate (UPR)*: The ratio of unclassified building pixels to the total number of building pixels;
- *Total Error Rate (TER)*: The ratio of the total number of wrongly classified pixels to the total number of pixels.

The results of the learned Dempster-Shafer classifier were compared with those of the Dempster-Shafer classifier without learning, i.e. with the initial probability assignment functions. In addition, the results of a less elaborated method of thresholding a Normalized DSM (nDSM) (Weidner and Forstner, 1995) were included in the comparison. This method is simply based on finding pixels in the *le-dtm* feature set whose values exceed a height threshold (e.g. 2 meters).

Table 2 summarizes the error measures obtained for the learned Dempster-Shafer classifier, the Dempster-Shafer method without learning, and the nDSM method. As can be seen, the learned Dempster-Shafer method yields a lower false positive rate, false negative rate and total error rate in comparison with the other two methods. The unclassified positive rates, however, show that the learned Dempster-Shafer classifier leaves about 5% of the building pixels unclassified. The rate of unclassified pixels can be interpreted as the rate of pixels for which the extracted features did not provide sufficient evidence for a correct classification. Figure 7 shows the distribution of the false positive, false negative and unclassified pixels. It is evident that the Dempster-Shafer method without learning misses a large number of building pixels, and with the nDSM method many trees are confused with buildings. The result of the Dempster-Shafer with learning shows a better performance where only a small number of building parts appear to be missed.

Method	False Negative Rate (FNR %)	False Positive Rate (FPR %)	Unclassified Positive Rate (UPR %)	Total Error Rate (TER %)
nDSM	5.51	10.69	0.00	8.76
Dempster-Shafer without learning	8.38	4.04	1.80	5.66
Dempster-Shafer with learning	4.40	3.35	5.14	3.74

Table 2: Results of the learned Dempster-Shafer classifier in comparison with those of the Dempster-Shafer method without learning and the method of thresholding a normalized DSM (nDSM).

A visualized comparison of the error metrics is provided in Figure 8. It can be noted that the Dempster-Shafer method without learning already outperforms the nDSM method in terms of the total error. Although the nDSM method has a lower false negative rate, the Dempster-Shafer without learning performs much better in terms of the false positive rate. The learning model improves the performance of the Dempster-Shafer classifier over all metrics. In particular, the improvement is more noticeable in the rate of false negative pixels, where the learned Dempster-Shafer model yields an error rate that is about two times smaller than that of the Dempster-Shafer without learning.

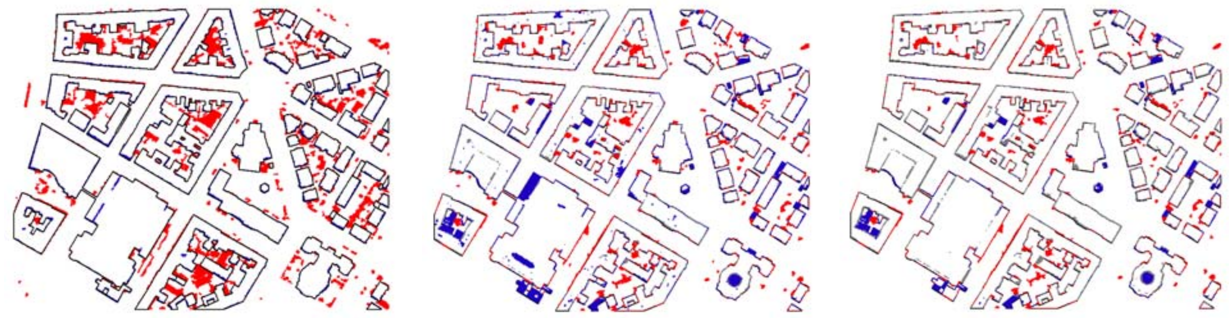

Figure 7: Distribution of false positive (red), false negative (blue) and unclassified positive (gray) pixels in the results of nDSM (left), Dempster-Shafer without learning (middle), and Dempster-Shafer with learning (right).

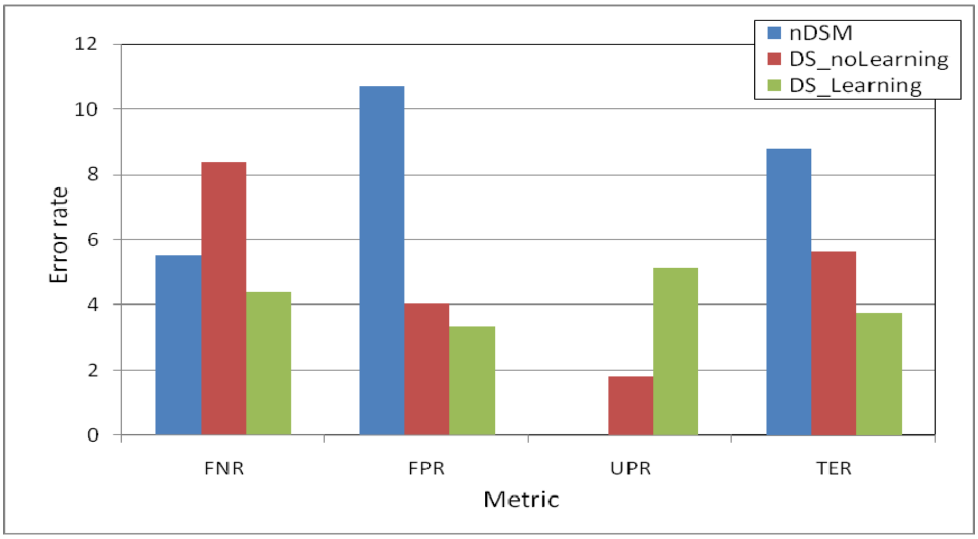

Figure 8: Comparison of the error metrics indicating the performance of the three methods.

5. Concluding Remarks

The process of evidence assignment in the Dempster-Shafer classifier usually requires interactive tuning of the mass probability assignment functions. In this paper, we presented a method for automated learning of the mass probability assignment functions from a set of training samples. In the context of building detection, the training samples can be automatically derived from an existing GIS database. Therefore, the learning Dempster-Shafer model provides a method for automated detection of buildings in multi-source aerial data. Our results show that the learning model improves the performance of the Dempster-Shafer classifier in detecting buildings in an urban study area.

The present implementation of the method works with three features extracted from image and range data. Future research will focus on extending the learning model with additional features including various texture descriptors. Another direction for future research is to investigate the relevance of features for a more accurate and reliable detection of buildings.

Acknowledgements

We would like to thank TopoSys GmbH for providing the dataset for our experiments.

References:

[1] Basir, O., Karray, F. and Zhu, H., 2005. Connectionist-based Dempster-Shafer evidential reasoning for data fusion. IEEE Transactions on Neural Networks, 16(6): 1513-1530.

[2] Binaghi, E., Gallo, I. and Madella, P., 2000. A neural model for fuzzy Dempster-Shafer classifiers. International Journal of Approximate Reasoning, 25(2): 89-121.

[3] Binaghi, E. and Madella, P., 1999. Fuzzy Dempster-Shafer reasoning for rule-based classifiers. International Journal of Intelligent Systems, 14(6): 559-583.

[4] Dempster, A.P., 1967. Upper and lower probabilities induced by a multi-valued mapping. The Annals of Mathematical Statistics, 38: 325-339.

[5] Frontoni, E., Khoshelham, K., Nardinocchi, C., Nedkov, S. and Zingaretti, P., 2008. Comparative analysis of automatic approaches to building detection from multi-source aerial data. In: G.J. Hay, T. Blaschke and D. Marceau (Editors), GEOBIA 2008 – Pixels, Objects, Intelligence. GEOgraphic Object Based Image Analysis for the 21st Century, Calgary, Canada.

[6] 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.

[7] Jang, J.S.R., 1993. ANFIS: Adaptive-Network-Based Fuzzy Inference System. IEEE Transactions on Systems Man and Cybernetics, 23(3): 665-685.

[8] Khoshelham, K., Nedkov, S. and Nardinocchi, C., 2008. A comparison of Bayesian and evidence-based fusion methods for automated building detection in aerial data. In: J. Chen, J. Jiang and H.-G. Maas (Editors), International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B7, Beijing, pp. 1183-1188.

[9] 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.

[10] Rottensteiner, F., Trinder, J., Clode, S. and Kubik, K., 2005. Using the Dempster–Shafer method for the fusion of LIDAR data and multi-spectral images for building detection. Information Fusion, 6(4): 283-300.

[11] Rottensteiner, F., Trinder, J., Clode, S., Kubik, K. and Lovell, B., 2004. 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.

[12] Shafer, G., 1976. A mathematical theory of evidence. Princeton University Press, Princeton, 292 pp.

[13] Sosnowski, Z. and Walijewski, J., 2005. Genetic-Based Tuning of Fuzzy Dempster-Shafer Model, Computational Intelligence, Theory and Applications. Springer, Berlin Heidelberg, pp. 747-755.

[14] Sugeno, M., 1985. Industrial applications of fuzzy control. Elsevier Science Inc., New York, NY, USA, 269 pp.

[15] 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.

[16] Zadeh, L.A., 1973. Outline of a new approach to the analysis of complex systems and decision processes. IEEE Transactions on Systems, Man and Cybernetics, 3(1): 28-44.