COMBINING STATISTICAL AND NEURAL CLASSIFIERS USING DEMPSTER-SHAFER THEORY OF EVIDENCE FOR IMPROVED BUILDING DETECTION

John Trinder, Mahmoud Salah
School of Surveying and Spatial Information Systems
The University of New South Wales
UNSW SYDNEY NSW 2052, Australia
(j.trinder, m.gomah)@unsw.edu.au
Phone +61 2 938 54197, Fax +61 2 9313 7493

Ahmed Shaker, Mahmoud Hamed, Ali Elsagheer
Dept. of Surveying, Faculty of Engineering Shoubra
Benha University
108 Shoubra Street, Cairo, Egypt
ahmshaker@link.net, prof.mahmoudhamed@ yahoo.com,
alielsagheer@yahoo.com
Phone +2 2 2201 3257, Fax +2 2 2201 3257

Abstract

This paper describes an approach for building detection from multispectral aerial images and lidar data by combining the results derived from statistical and neural network classifiers, which offer complementary information, based on Dempster-Shafer Theory of Evidence. Four study areas with different sensors and scene characteristics were used. First, we filtered the lidar point clouds to generate a Digital Terrain Model (DTM), and then the Digital Surface Model (DSM) and the Normalised Digital Surface Model (nDSM) were generated. After that a total of 25 uncorrelated feature attributes have been generated from the aerial images, the lidar intensity image, DSM and nDSM. Then, three different classification algorithms were used to detect buildings from aerial images, lidar data and the generated attributes. The classifiers used include: Self-Organizing Map (SOM); Classification Trees (CTs); and Support Vector Machines (SVMs). The Dempster-Shafer theory of evidence was then applied for combining measures of evidence from the three classifiers. A considerable amount of the misclassified building pixels were recovered by the combination process.

Introduction

Research on building detection from aerial images and lidar data fusion has been undertaken so that the strengths of each data type can compensate for the weaknesses of the other. Low contrast, occlusions and shadow effects in the images can be compensated by the accurately defined planes in the lidar data. On the other hand, the poorly defined edges in the lidar data can be compensated by the accurately defined edges in the aerial images. Matikainen et al. (2007) applied the Gini splitting criterion for CT for building detection from aerial images and lidar data. Rottensteiner et al. (2007) evaluated the Dempster-Shafer based fusion of multispectral aerial images and lidar data for

building detection. Salah et al. (2009) tested the Self-Organizing Map (SOM) for building detection from aerial images and lidar data.

Kanellopoulos et al. (1997) have demonstrated the complementary behaviours of neural and statistical algorithms in terms of classification errors. The efficient combination of such classifiers, should achieve better classification results than any single classifier, even results obtained with the best classifier used individually. In the simplest implementation of the hybrid concept, predictions of different classifiers were averaged (Breiman, 1996). Many other methods have also been used to combine classifiers, such as Bagging (Breiman, 1996) and Boosting (Freund and Schapire, 1997). Applications of majority voting (MV) for pattern recognition have already been studied in detail in Lam and Suen (1997). Another technique which is widely studied in classical classifier fusion but less addressed in remote sensing is Dempster-Shafer (*D-S*) theory (Shafer, 1976). *D-S* theory has already been investigated in handwriting recognition, in automatic disambiguation of word senses, in human speech perception but this is probably the first attempt to use it for combining information derived from different classifiers for improvement of land cover mapping.

After introducing *D-S* theory in the following section, the methods are described, and then the results are presented and evaluated. Finally, the results are summarised.

Overview of Dempster-Shafer Theory

The theory of evidence was introduced by Shafer (1976) as a mathematical framework for representation and combination of different measures of evidence. It can be considered as a generalization of the Bayesian framework and permits the characterization of uncertainty and ignorance. In outlining the Dempster-Shafer theory, we consider a classification problem where the input data are to be classified into n classes $C_j \in \theta$, θ is referred to as *the frame of discernment*. The power set of θ is denoted by 2^θ i.e. the set of all subsets of θ. A probability mass $m(A)$ is assigned to every class $A \in 2^\theta$ by a classifier such that $m(\emptyset) = 0$, $0 \le m(A) \le 1$, and $\Sigma\, m(A) = 1$, where the sum is to be taken over all $A \in 2^\theta$ and $\emptyset$ denotes the empty set. $m(A)$ can be interpreted as the amount of belief that is assigned exactly to A and not to any of its subsets. 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 all probability masses not assigned to the complementary hypothesis $\bar{A}$ of A with $A \cap \bar{A} = \emptyset$ and $A \cup \bar{A} = \theta$:

$$Sup(A) = \sum_{B \subseteq A} m(B); \qquad Pls(A) = \sum_{A \cap B \neq \phi} m(B) = 1 - Sup(\bar{A}) \tag{1}$$

$Sup(A)$ is also called *dubiety*. It represents the degree to which the evidence contradicts a proposition. If z classes are available, probability masses $m_i(B_j)$ have to be defined for all these classes i with $1 \le i \le z$ and $B_j \in 2^\theta$. From these probability masses, a combined probability mass can be computed for each class $A \in 2^\theta$ through an *orthogonal summation* process as follow:

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

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. It is important to mention that the combination rule given by equation 2 assumes that the belief functions to be combined are independent.

Study Area and Data Sources

Four test datasets of different sensor and scene characteristics were used in this study as shown in table 1 and figure 1. These scenes include data available from data providers in Australia to the researchers together with that provided by TopoSys in Germany for a range of land covers, including high and low density urban areas, a rural township and a densely populated European town.

Table 1. Characteristics of image and lidar data sets.

Test area	Size	Lidar Data		Aerial images	
		Sensor	wavelength	bands	pixel size
UNSW	0.5 x 0.5Km	Optech ALTM 1225	1.047µm	RGB	10cm
Bathurst	1 x 1Km	Leica ALS50	1.064µm	RGB	50cm
Fairfield	2 x 2Km	Optech ALTM 3025	1.047µm	RGB	15cm
Memmingen	2 x 2Km	TopoSys	1.56µm	CIR	50cm

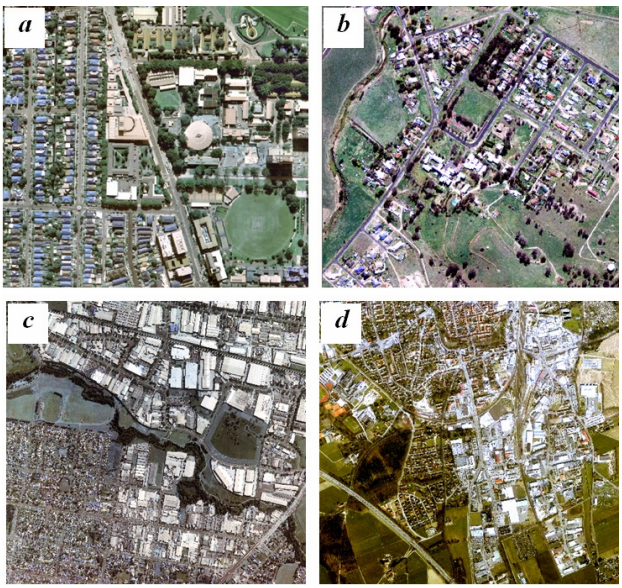

Figure 1. Orthophotos for: (a) UNSW; (b) Bathurst; (c) Fairfield; and (d) Memmingen.

In order to evaluate the accuracy of the results, reference data were captured by digitising buildings, trees, roads and ground in the orthophotos. Class "ground" mainly corresponds to grass, parking lots and bare fields. Larger areas covered by trees were digitised as one polygon. Information on single trees was captured where possible.

Methodology

Data Pre-Processing

First the original lidar point clouds were filtered to separate on-terrain points from points falling onto natural and human made objects. A filtering technique based on a linear first-order equation which describes a tilted plane surface has been used (Salah et al., 2009). Data from both the first and the last pulse echoes were used in order to obtain denser terrain data and hence a more accurate filtering process. After that, the filtered lidar points were converted into an image Digital Terrain Model (DTM), and the Digital Surface Model (DSM) was generated from the original lidar point clouds. Then, the Normalized Digital Surface model (nDSM) was generated by subtracting the DTM from the DSM. Finally, a height threshold of 3m was applied to the nDSM to eliminating other objects such as cars to ensure that they are not included in the final classified image. Our experiments were carried out characterizing each pixel by a 32-element feature vector which comprise: 25 generated attributes, 3 image bands (R, G and B), intensity image, DTM, DSM and nDSM. The 25 attributes include those derived from the Grey-Level Co-occurrence Matrix (GLCM), Normalized Difference Vegetation Indices (NDVI), slope and the polymorphic texture strength based on the Förstner operator (Förstner and Gülch, 1987). The attributes were calculated for pixels as input data for the three classifiers, and were selected to be uncorrelated. Table 2 shows the attributes and the images for which they have been derived. We selected 1644, 1264, 1395 and 1305 training pixels for buildings, trees, roads and ground respectively for each band of the input data. Class “ground” mainly corresponds to grass, parking lots and bare fields. A detailed description of the filtering and generation of attributes process can be found in Salah *et al.* (2009).

Table2. The full set of the possible attributes from aerial images and lidar data. √ and x indicate whether or not the attribute has been generated for the image.

Attribute	Red Band	Green Band	Blue Band	Intensity/IR	DSM	NDSM
Polymorphic strength	√	√	√	√	√	√
GLCM/homogeneity	√	√	√	√	√	√
GLCM/mean	√	√	√	√	√	√
GLCM/entropy	√	√	√	√	√	√
Slope	x	x	x	x	x	√

Land Cover Classification

In this work, three classifiers have been used to estimate the class memberships required for the combination process as follows:

Support Vector Machines (SVMs)

SVMs are based on the principles of statistical learning theory (Vapnik, 1979). SVMs delineate two classes by fitting an optimal separating hyperplane (OSH) to those training samples that describe the edges of the class distribution. As a consequence they generalize well and often outperform other algorithms in terms of classification accuracies. The One-Against-All (1AA) technique was used to solve for the binary classification problem that exists with the SVMs and

to handle the multi-class problems in aerial and lidar data. The Gaussian radial basis function (RBF) kernel has been used, since it has proved to be effective with reasonable processing times in remote sensing applications. In order to specify the RBF parameters (the penalty parameter, *C*, and the width of the kernel function, *γ*), a grid-search on *C* and *γ* using a 10-fold cross-validation was used. The original output of a SVM represents the distances of each pixel to the optimal separating hyperplane, referred to as rule images. All positive (+1) and negative (-1) votes for a specific class were summed and the final class membership of a certain pixel was derived by a simple majority voting.

Self-Organizing Map Classifier (SOM)

The SOM (Kohonen, 2001), as a neural network classifier, requires no assumption regarding the statistical distribution of the input pattern classes and has two important properties: the ability to *learn* from input data; and to generalize and predict unseen patterns based on the data source. In this work (Salah et al., 2009), the SOM has 32 input neurons which are: 25 generated attributes, 3 image bands (R, G and B), intensity image, DTM, DSM and nDSM. The output layer of an SOM was organized as a 15 x 15 array of neurons as an output for the SOM (255 neurons). This number was selected because small networks result in some unrepresented classes in the final labelled network, while large networks lead to an improvement in the overall classification accuracy. Initial synaptic weights between the output and input neurons were randomly assigned (0-1). In the output of the SOM, each pixel is associated with a degree of membership for a certain class.

Classification Trees (CTs)

The theory of Classification Trees (CTs) (also called decision trees) was developed by Breiman et al. (1984). This is an iterative procedure in which a heterogeneous set of training data consisting of multiple classes is hierarchically subdivided progressively into more homogeneous clusters using a binary splitting rule to form the tree, which is then used to classify other similar datasets. In the final classification not all but only the most prominent attributes are used. The Entropy model was used as the splitting criteria in our study. Also, the trees were pruned through a 10-fold cross validation process. In the original output of the CTs, each pixel is associated with a degree of membership for the class at which particular leaf it was classified.

The proposed approach for Dempster-Shafer Based Combination

First, we considered the output from each classifier as a source of information. After that, Dempster-Shafer theory was run to combine the outputs of the different classifiers. The class membership estimates at the output of each classifier (pp_i) were weighted according to the classifiers' reliability, accuracy rate, for each class ($0 \leq \alpha_{ci} \leq 1$) to solve the problem when a classifier gives equal class memberships for all the classes at a certain pixel. This results in new basic probability assignment (*BPAs*), *m*, defined by:

$$m_i(A_i) = \alpha_{ci}\ pp_i(A_i) \tag{3}$$

$$m_i(\bar{A}_i) = \alpha_{ci} \sum_{j \neq i} pp(A_j) \tag{4}$$

After scaling down the mass functions, if the summation of $m_i(A_i)$ and $m_i(\overline{A}_i)$ *was* less than 1, the remainder was assigned to the frame of discernment as ignorance. This remainder indicates that the classifier was not able to decide on an appropriate class, so it was interpreted as lack of information according to *D-S* theory.

$$m_i(\theta_i) = 1 - m_i(A_i) - m_i(\overline{A}_i) \quad (5)$$

These *BPAs* are then combined making use of Dempster-Shafer's rule of combination to obtain the final combined *BPAs*. The procedure produces a set of belief images (one per class) showing the degree of belief that each pixel belongs to each class. These images were converted into land cover classified maps by selecting the class image that contains the maximum belief and assigning that class to the output pixel.

Results and Discussion

Evaluation of the Proposed Method

The overall classification accuracies of individual classifiers, based on the reference data, were evaluated first with the overall accuracy of the best classifier serving as a reference. Several of the most widely used probability combination strategies were also tested and compared to our proposed method. These strategies include: Maximum Rule (*MR*); Fuzzy Majority Voting (*FMV*) using the relative quantifier at least half with the parameter pair (0, 0.5); and Weighted Sum (*WS*). A detailed description of these combination methods can be found in Yager (1998). The overall classification accuracies of individual classifiers, based on the reference data, are given in table 3.

Table 3. Performance evaluation of single classifiers for the four test areas.

	Classification accuracy (%)		
Test area	SOM	CT	SVM
UNSW	96.8	95.05	96.9
Bathurst	95	92.85	96.5
Fairfield	96.8	96.15	97
Memmingen	95	90.75	96.6
Mean	95.9	93.7	96.75
SD	1.04	2.40	0.24

The improvement in overall classification accuracies achieved by each combination method compared with the best individual classifier, SVM, was determined as shown in figure 2. For the four test areas, it is clear that the overall performances of *D-S* are better than those of the other combination methods. *D-S* performs slightly better than *FMV*. *MR* resulted in the worst performance followed by the *WS*. The question still remains as to whether these improvements are statistically significant. To answer this question, first, the standard deviation (SD) of the classification accuracies produced by each classifier for the four test areas is determined to express the variability in classification accuracies from the mean as shown in table 3. The low standard deviation of 0.24% for the SVM results indicates that the spread of the accuracies for the four tests areas is small and hence accuracies tend to be very close to the mean. We can assume that the reported margin of error is

typically about plus/minus twice the standard deviation (a range for an approximately 95 percent confidence interval). For this work we used a margin of accuracy of 0.72%, which is three times the standard deviation of the SVMs results, to define the improvements in accuracy that are considered statistically significant, as shown by the dashed line in figure 2. Any improvements in classification accuracy more than the dashed horizontal line are deemed to be significant. It can be concluded that the application of *D-S* results in the most significant improvement in classification accuracy. The improvements achieved by other techniques are either extremely close to the significance value, and therefore considered to be marginally significant, or below the value of significance. Taking into account the limited room for improvement beyond 96.9% accuracy caused by other errors in image acquisition and image to lidar geographic registration, the best average improvement in classification accuracy of 1.1% is obtained from *D-S* algorithm. *FMV* algorithm gives an average improvement of 0.9%. *MR* resulted in the worst performance and only improved the results by 0.7% followed by *WS* with 0.8% average improvement.

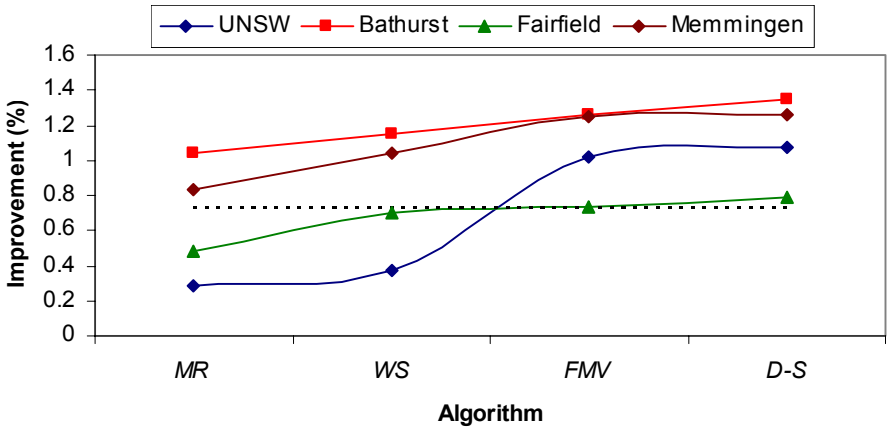

Figure 2. Performance comparison of the proposed fusion algorithms with existing fusion algorithms, compared with the performance of the best individual classifier, SVM. Improvements exceeding the dashed horizontal line are considered to be significant.

Performance Analysis of the Proposed Method for Individual Classes

Two additional measures were used to evaluate the performance of the proposed combination method, namely: commission and omission errors. Unlike overall classification accuracy, commission and omission errors clearly show how the performance of the proposed method improves or deteriorates for each individual class in the combined classifiers. Table 4 shows the commission and omission errors based on *D-S* combination, compared to the commission and omission errors of the best individual classifier in case of UNSW. It can be seen that most of the class-commission and omission errors are reduced by the *D-S* fusion. Contrarily, there was an increase in commission and/or omission errors for a few classes. However, those classes are still classified with relatively low commission and omission errors. Another advantage of the *D-S* fusion over SVMs is that the achieved errors are less variable. The application of *D-S* fusion significantly reduced the SD for commission and omission errors. The visual assessment interpretation, figure 3, clearly shows a relatively high degree of noise in the SVMs-based classification results. In contrast to this, the classification that is based on the *D-S* appears more homogenous.

Table 4. Comparison of errors using the best classifier, SVM, with the classification resulting from the D-S combination. B, T, R and G refer to buildings, trees, roads and grass respectively.

		Best classifier (SVMs)		*DS fusion*	
		Commission (%)	Omission (%)	Commission (%)	Omission (%)
UNSW	B	4.65	2.77	1.50	0.95
	T	3.18	1.97	2.15	1.37
	R	4.81	0.06	0.12	1.51
	G	0.06	5.10	1.55	0.12
Bathurst	B	9.79	7.80	5.98	3.57
	T	0.35	6.12	0.90	0.08
	R	4.36	0.98	1.62	4.07
	G	10.30	4.06	0.40	0.57
Fairfield	B	8.23	11.11	0.99	2.79
	T	0.89	3.36	4.06	1.46
	R	4.08	0.76	0.11	1.61
	G	3.69	7.04	1.65	0.11
Memming.	B	4.04	21.28	1.50	0.95
	T	0.63	3.94	1.37	2.15
	R	4.10	0.42	1.51	0.12
	G	7.96	5.30	0.12	1.55
Mean		4.45	5.13	1.60	1.44
SD		3.22	5.25	1.52	1.22

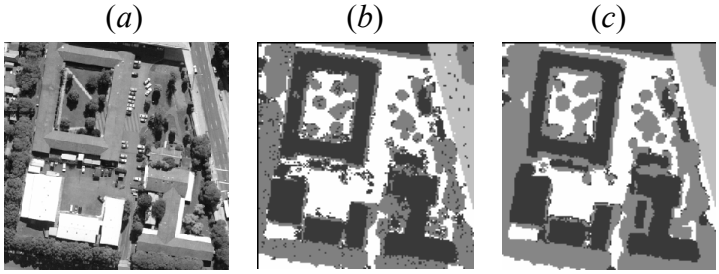

Figure 3. (a) Aerial image; (b) classification results of the best classifier (SVMs); (c) Error correction after applying the D-S fusion algorithm. Black: buildings; dark grey: trees; light grey: roads; white: ground.

Completeness and correctness of the extracted buildings

In order to extract buildings from the classified data, the classified image is converted to a binary image, with pixels representing buildings displayed with one, and non-building pixels (roads, trees and ground) as zeros. Then, the smaller raster homogeneous building regions were merged into the larger neighbouring homogeneous regions or deleted according to an arbitrary 1m distance and 30m^2 area threshold. The area threshold represents the expected minimum building area that can be reliably extracted, while the distance threshold was set to 1m to fill in any gaps produced through the classification process. As a last step, building borders were cleaned by removing small structures that were connected to building borders. The result was a black and white image that represents the detected buildings without noisy features and also without holes as shown in figure 4/left. In order to evaluate the performance of the building extraction process, the *completeness* and the *correctness* of the detected buildings were investigated based a per-building level. A building in the reference data set is counted as a true positive if at least 80% of its area is covered by buildings detected automatically and vice versa.

Figure 4/right shows the completeness and correctness against the building size for UNSW case study for SOM, CTs, SVMs classifiers as well as the combined classifier using the *D-S* as a typical example obtained for these experiments. For the UNSW case study, buildings around 30m^2 were detected with average improvements in completeness and correctness over the best classifier of around 0.3% and 0.7% respectively. For Bathurst, Fairfield and Memmingen case studies, buildings around 30m^2 were detected with average improvements in completeness and correctness over the best classifier of around: (1.4%,1.5%); (0.75%, 0.7%); and (1.2%, 1.1%) respectively. For all cases, all buildings larger than 60m^2 were detected with average improvements in completeness and correctness over the best classifier of around 1.9%. These tests strongly represent achievable accuracies for detection of buildings by the proposed method using a combination of lidar data and images.

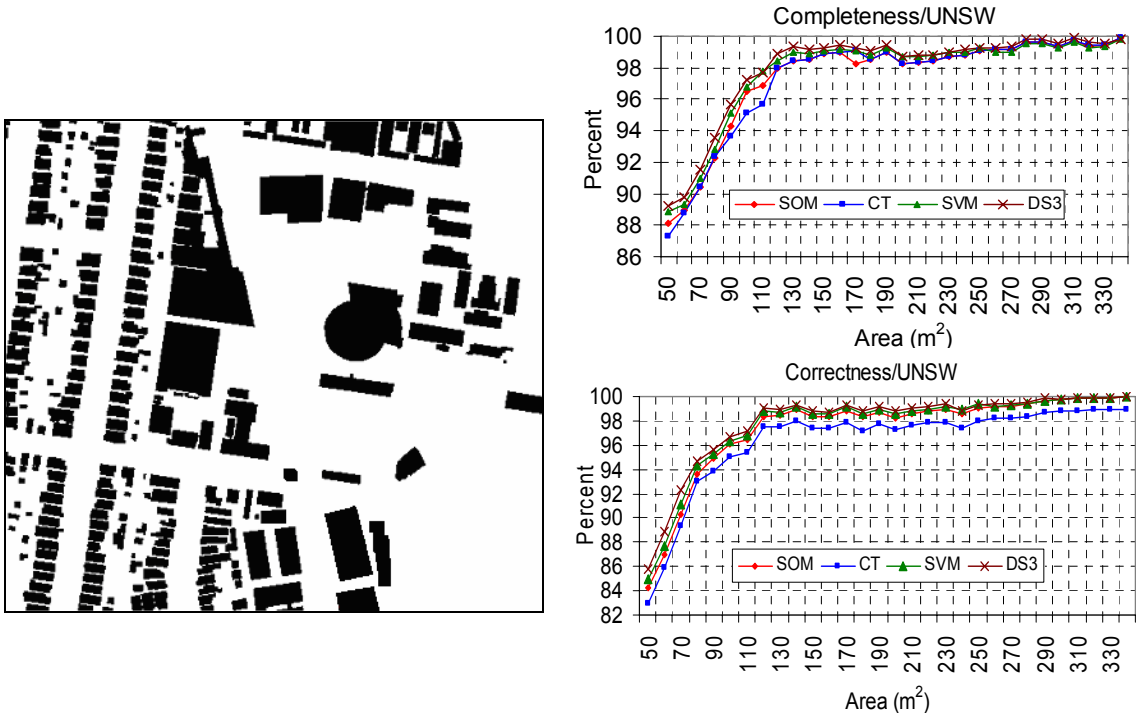

Figure 4. Left:The final 2-D binary image for UNSW test area; right: Completeness and correctness derived for SOM, CTs, SVMs and combined classifier using the D-S algorithm plotted against building areas in cases of UNSW.

Conclusion

In this paper, we developed a new powerful Multiple Classifier System (MCS) to combine statistical and neural classifiers based on the *D-S* theory of evidence. The system first weights the class membership at each classifier output based on the reliability of that classifier, then it determines the amount of belief from a given classifier output that must be discarded i.e. assigned to the ignorance hypothesis. The results showed an improvement in terms of overall classification accuracy and omission and commission errors of individual classes. The proposed fusion algorithm gives an accuracy of 98%, which is an improvement of around 1.1% over the best classifier. This is an enhancement considering the limited room for improvement beyond 96.9% accuracy achieved with the SVM classifier. On the other hand, the average commission and

omission errors have been reduced by about 64% and 72% respectively compared to the best single classifier. A comparison of the results with some of the existing fusion rules demonstrates that the proposed fusion algorithm gives the best results.

References

BREIMAN, L., 1996, Bagging predictors, *Machine Learning*, 24(2):123-140.

BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A. and STONE, C. J., (editors), 1984, *Classification and Regression Trees*, Chapman & Hall, New York, 358 p.

FÖRSTNER, W., and GÜLCH, E., 1987, A fast operator for detection and precise location of distinct points, Corners and Centres of Circular Features. *Proceedings of the ISPRS 1987 Intercommission Workshop on Fast Processing of Photogrammetric Data*, 2-4 June 1987, Interlaken, Switzerland, pp. 281-305.

FREUND, Y., and SCHAPIRE, R.E., 1997, A decision-theoretic generalization of online learning and application to boosting, *Journal of Computer and System Science*, 55(1):119-139.

KANELLOPOULOS, I., WILKINSON, G., ROLI, F. and AUSTIN, J., (editors), 1997, *Neurocomputation in Remote Sensing Data Analysis*, Springer, Berlin.

KOHONEN, T.,2001, *Self-Organizing Maps*. Third Edition, Springer, New York.

Lam, L., and CY. Suen, 1997, Application of majority voting to pattern recognition: an analysis of its behaviour and performance. *IEEE Transactions on Systems*, *Man, and Cybernetics*, 27(5): 553-568.

MATIKAINEN, L., KAARTINEN, H. and HYYPPÄ, J., 2007, Classification tree based building detection from laser scanner and aerial image data. In *Proceedings of the ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007*, 12–14 September 2207, Espoo, Finland, pp. 280–287.

ROTTENSTEINER, F., TRINDER, J., CLODE, S. and KUBIK, K., 2007, Building detection by fusion of airborne laser scanner data and multi-spectral images: Performance Evaluation and Sensitivity Analysis, *ISPRS Journal of Photogrammetry & Remote Sensing*, 62, pp. 135–149.

SALAH, M., TRINDER, J. and SHAKER, A., 2009, Evaluation of the self-organizing map classifier for building detection from lidar data and multispectral aerial images. *Journal of Spatial Science*, 54:15-34.

SHAFER, G., 1976, A mathematical theory of evidence. Princeton University Press, 297 p.

VAPNIK, V., 1979, *Estimation of Dependences Based on Empirical Data* [in Russian]. Nauka, Moscow, 1979. (English translation: Springer Verlag, New York, 1982).

YAGER, R.R., 1998, On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Transactions on Systems, Man, and Cybernetics*, 18: 183-190.