

INTEGRATING MULTIPLE CLASSIFIERS WITH FUZZY MAJORITY VOTING FOR IMPROVED LAND COVER CLASSIFICATION

M. Salah^{a,*}, J.C.Trinder^a, A.Shaker^b, M.Hamed^b, A.Elsagheer^b

^a School of Surveying and Spatial Information Systems, The University of New South Wales, UNSW SYDNEY NSW 2052, Australia - (m.gomah, j.trinder)@unsw.edu.au

^b Dept. of Surveying, Faculty of Engineering Shoubra, Benha University, 108 Shoubra Street, Cairo, Egypt - ahmshaker@link.net, prof.mahmoudhamed@yahoo.com, alielsagheer@yahoo.com

ISPRS WG III/2 "Point Cloud Processing"

KEY WORDS: Aerial Images, Lidar, GLCM, Attributes, Hybrid Classifier, Fuzzy Majority Voting.

ABSTRACT:

In this paper the idea is to combine classifiers with different error types based on Fuzzy Majority Voting (FMV). Four study areas with different sensors and scene characteristics were used to assess the performance of the model. First, the lidar point clouds were filtered to generate a Digital Terrain Model (DTM), and then a Digital Surface Model (DSM) and the Normalized Digital Surface Model (nDSM) were generated. A total of 25 uncorrelated feature attributes have been generated from the aerial images, the lidar intensity image, DSM and nDSM. Three different classification algorithms were used to classify buildings, trees, roads and ground from aerial images, lidar data and the generated attributes. The used classifiers include: Self-Organizing Map (SOM); Classification Trees (CTs); and Support Vector Machines (SVMs) with average classification accuracies of 96.8%, 95.9% and 93.7% obtained for SVMs, SOM, and CTs respectively. FMV was then applied for combining the class memberships from the three classifiers. The main aim is to reduce overlapping regions of different classes for minimizing misclassification errors. The outcomes demonstrate that the overall accuracy as well as commission and omission errors have been improved compared to the best single classifier.

1. INTRODUCTION

Researchers are continually seeking to improve the performance of classifiers in remote sensing. Taking advantage of the complementary information about image data provided by classifiers based on different mathematical concepts, the next natural frontier is the integration of multiple approaches into a unified framework. The efficient combination of classifiers, should achieve better classification results than any single classifier. Kanellopoulos et al. (1997) have demonstrated the complementary behaviours of neural and statistical algorithms in terms of classification errors. Therefore these classifiers result in uncorrelated classification errors and hence higher accuracies can then be reached by combining them.

2. RELATED WORK

For remote sensing applications, Benediktsson et al. (2007) have presented a brief summary of recent developments of multiple classifier systems (MCS) in which the optimal set of classifiers is first selected and then they are combined by a specific fusion method. The aim is to effectively merge the results of the classifiers taking advantage of the benefits of each while reducing their weaknesses.

More recently, researchers have investigated classifier selection for MCS design. Giacinto and Roli (2001) clustered the candidate classifiers according to interdependency and selected one classifier from each cluster. Hao et al. (2003) also used a heuristic search for classifier selection. Mountrakis et al. (2009) presented a hierarchical, multi-stage adaptive strategy for image classification. They iteratively applied various classification methods, e.g., decision trees, neural networks, identified regions

of parametric and geographic space where accuracy is low, and tested the application of alternate methods, repeating the process until the entire image was classified.

Applications of majority voting (MV) for pattern recognition have already been studied in detail in Lam and Suen (1997). A trainable variant of majority voting is weighted majority voting, which applies a weight to each vote. The weight applied to each classifier can be obtained for example by estimating the accuracies of the classifiers on a validation set.

Yu-Chang and Kun-Shan (2009) introduced a multiple classifier system for land cover classification. The Bagging and Boosting algorithms were investigated as a weighting policy and then an adaptive thresholding criterion was defined to account for the ambiguities between classes.

Recent work has focused on deriving the uncertainty map of the land-cover prediction, which based on the uncertainty of land-cover classification for each pixel. Alimohammadi et al. (2004) used maximum likelihood classification algorithm to perform the classification and generated uncertainty estimation.

Another technique which is widely studied in classical classifier fusion but less addressed in remote sensing is Fuzzy Majority Voting (FMV). FMV theory has already been investigated in automatic disambiguation of word senses (Le et al., 2007), but this is probably the first attempt to use it for combining information derived from different classifiers for improvement of land cover mapping. FMV has been proposed in this research to further improve the classification performances and overcome the shortcomings of the previous approaches of combining classifiers, such as sensitivity to noise, computational load and the need for parametric statistical modeling of each data source. The major motivation of our work is to establish a framework to combine classifiers with

* Corresponding author.

different error types based on FMV, thus minimizing the misclassification errors. After describing the study areas and data sources in the following section, this paper is organised as follows. Section 4 describes the methods. Section 5 presents and evaluates the results and we summarise our results in Section 6.

3. STUDY AREAS AND DATA SOURCES

3.1 Test Zones and Input Data

Four test datasets of different sensor and scene characteristics were used in this study as summarized in Table 1 and 2. Test area 1 is a part of the region surrounding the University of New South Wales campus, Sydney Australia, which is a largely urban area that contains residential buildings, large Campus buildings, a network of main roads as well as minor roads, trees, open areas and green areas. The colour imagery was captured by film camera at a scale of 1:6000. The film was scanned in three colour bands (red, green and blue) in TIFF format, with 15µm pixel size (GSD of 0.09m) and radiometric resolution of 16-bit as shown in figure 1(a). Test area 2 is a part of Bathurst city, NSW Australia, which is a largely rural area that contains small residential buildings, road networks, trees and green areas. The colour (red, green and blue) images were captured by a Leica ADS40 line scanner sensor and supplied as an ortho image as shown in figure 1(b). Test area 3 is over suburban Fairfield, NSW Australia covering low density development in the southwest half of the scene, and large industrial buildings in the northeast part as shown in figure 1(c). The image data was acquired by a film camera at a scale of 1:10 000 which was scan digitized and supplied as an ortho image. Test area 4 is over Memmingen Germany, featuring a densely developed historic centre in the north of the scene and industrial areas in the remainder as shown in figure 1(d). Multispectral images (CIR), including an infrared image with the same resolution as the colour bands, were acquired by a line scanner sensor and supplied as an ortho image.

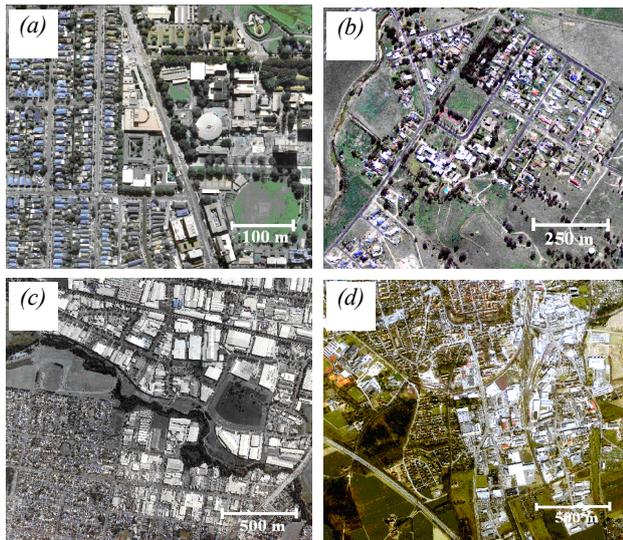


Figure 1. Orthophotos for: (a) UNSW; (b) Bathurst; (c) Fairfield; and (d) Memmingen.

Test area	Size (Km)	bands	pixel size (cm)	Camera
UNSW	0.5 x 0.5	RGB	9	LMK1000
Bathurst	1 x 1	RGB	50	ADS40
Fairfield	2 x 2	RGB	15	Line scanner
Memmingen	2 x 2	CIR	50	LMK1000
				TopoSys Falcon II line scanner

Table 1. Characteristics of image datasets.

	UNSW	Bathurst	Fairfield	Memmingen
	Optech ALTM 1225	Leica ALS50	Optech ALTM 3025	TopoSys
Spacing across track (m)	1.15	0.85	1.2	0.15
Spacing along track (m)	1.15	1.48	1.2	1.5
Vertical accuracy (m)	0.10	0.10	0.15	0.15
Horizontal accuracy (m)	0.5	0.5	0.5	0.5
Density (Points/m ²)	1	2.5	1	4
Sampling intensity (mHz)	11	150	167	125
Wavelength (µm)	1.047	1.064	1.047	1.56
Laser swath width (m)	800	777.5	700	750
Recorded pulse	1 st and last			

Table 2. Characteristics of lidar datasets.

3.2 Training Datasets

All tests were conducted using identical training sets. Eighty polygons of approximately equal areas, twenty for each land cover class, buildings, trees, roads and ground, were overlaid over each image to generate the training data. The positions of the polygons were selected carefully to be representative and to capture changes in the spectral variability of each class. The training data for each test area consists of 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.

3.3 Reference Data

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. We chose to digitize all recognisable features independently of their size. Adjacent buildings that were joined but obviously separated were digitized as individual buildings. Otherwise, they were merged as one polygon. In order to overcome the horizontal layover problem of tall objects such as buildings, roofs were first digitized and then each roof polygon was shifted if possible so that at least one point of the polygon coincided with the corresponding point on the ground. For Fairfield, the orthophoto and the lidar data correspond to different dates. Thus, we excluded from the analysis 41 building polygons that were only available in one data set. Larger areas covered by trees were digitised as one polygon. Information on single trees was captured where possible.

4. METHODOLOGY

The combination process was implemented in several stages as follow:

4.1 Filtering of lidar point clouds

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 DTM, and the DSM was generated from the original lidar point clouds. Then, the 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.

4.2 Generation of Attributes

Our experiments were carried out characterizing each pixel by a 32-element feature vector which comprises: 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 NDVI values for the UNSW, Bathurst and Fairfield test areas were derived from the red image and the lidar reflectance values, since the radiation emitted by the lidars is in the IR wavelengths. The resolutions of the lidar reflectance data for these study areas are lower than that for the images, and this may impact on the ability to detect vegetation. Since the images derived for the Memmingen dataset include an IR channel, the NDVI was derived from the image data only. The attributes were calculated for pixels as input data for the three classifiers. Table 3 shows the attributes and the images for which they have been derived. These attributes have been selected to be uncorrelated based on the problem of correlation between feature attributes. All the presented attributes were used for every test area. A detailed description of the filtering and generation of attributes process can be found in Salah et al. (2009).

attribute	Red Band	Green Band	Blue Band	Intensity	DSM	nDSM
PTS	√	√	√	√	√	√
HMG	√	√	√	√	√	√
Mean	√	√	√	√	√	√
entropy	√	√	√	√	√	√
Slope	x	x	x	x	x	√

Table 3. 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. PTS refers to polymorphic texture strength; HMG refers to GLCM/homogeneity; Mean refers to GLCM/ Mean; entropy refers to GLCM/ entropy.

4.3 Land Cover Classification

In this work, we have used the Self-Organizing Map (SOM), Classification Trees (CTs), and Support Vector Machines

(SVMs) classifiers to estimate the class memberships required for the combination process.

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. Furthermore, the misclassification errors are minimized by maximizing the margin between the data points and the decision boundary. Since the One-Against-One (1A1) technique usually results in a larger number of binary SVMs and then in subsequently intensive computations, 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. Two parameters should be specified while using RBF kernels:

- C , the penalty parameter that controls the trade-off between the maximization of the margin between the training data vectors and the decision boundary plus the penalization of training errors
- γ , the width of the kernel function.

In order to estimate these values and to avoid making exhaustive parameter searches by approximations or heuristics, we used a grid-search on C and γ using a 10-fold cross-validation. 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 undertakes both unsupervised and supervised classification of imagery using Kohonen's SOM neural network (Kohonen, 2001). SOM 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, rather than on any particular *a priori* model. 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 (225 neurons). This number was selected because, as recommended by Hugo et al. (2006), 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). A CT is a non-parametric univariate technique built through a process known as binary recursive partitioning. 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. CTs have the advantage that they also work when the classification variables are a mixture of categorical and continuous. In the final classification not all but only the most prominent attributes are used. This makes the classification method highly automatic and different from most other approaches in which the input data must remain fixed. The Entropy model was used as the splitting criteria in our study. Also, the trees were pruned through a 10-fold cross validation process, which has been demonstrated to produce highly accurate results without requiring an independent dataset for assessing the accuracy of the model. 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. If a pixel is not associated with that class, it will be assigned a zero.

4.4 Fuzzy Majority Voting Based Combination

The idea is to give some semantics or meaning to the weights. Therefore, based on these semantics the values for the weights can be provided directly. In the following the semantics based on fuzzy linguistic quantifiers for the weights are used. The fuzzy linguistic quantifiers were introduced by Zadeh (1983), who defined two basic types of quantifiers: absolute, and relative. Here the focus is on relative quantifiers typified by terms such as ‘most’, ‘at least half’, or ‘as many as possible’. The membership function of relative quantifiers for a given pixel as given by the i^{th} classifier can be defined as (Herrera and Verdegay, 1996):

$$Q_{pp_i} = \begin{cases} 0 & \text{if } pp_i < a \\ \frac{pp_i - a}{b - a} & \text{if } a \leq pp_i \leq b \\ 1 & \text{if } pp_i > b \end{cases} \quad (1)$$

With parameters $a, b \in [0, 1]$, and pp_i is the class membership of the pixel as obtained for the i^{th} classifier. Then, Yager (1988) proposed to compute the weights based on the linguistic quantifier represented as follows:

$$w_{pp_i} = Q_{pp_i} \left(\frac{J_i}{N} \right) - Q_{pp_i} \left(\frac{J_i - 1}{N} \right) \quad (2)$$

Q_{pp_i} is the membership functions of relative quantifiers for the pixel as obtained for the i^{th} classifier. J_i is the order of the i^{th} classifier after ranking Q_{pp_i} values of the pixel, for all classifiers, in a descending order. N is the total number of classifiers.

The relative quantifier ‘at least half’ with the parameter pair (0, 0.5) for the membership function Q_{PP} in equation 1 as graphically depicted in figure 2 was used. Depending on a particular number of classifiers N , 3 in our case, and by using equation 2, the corresponding weighting vector of the given pixel, $W_{PP} = [w_{PP1}, \dots, w_{PPN}]$ can be obtained. Finally, the probability based on FMV (P_{FMV}) can be calculated as follows:

$$P_{FMV} = \arg \max_k \left[\sum_{i=1}^N w_{pp_i} pp_i \right] \quad (3)$$

with k is the number of classes.

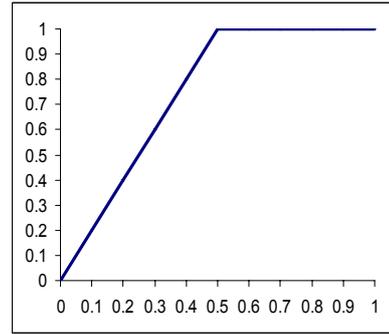


Figure 2. Linguistic Quantifier at least half with the parameter pair (0, 0.5).

4.5 Evaluation of the Proposed Method

The overall classification accuracies of individual classifiers, based on the reference data, were evaluated first and the overall accuracy of the best classifier served as a reference. Two of the most widely used probability combination strategies were also tested and compared to the proposed method. These strategies include: Maximum Rule (MR); and Weighted Sum (WS). A detailed description of these combination methods can be found in Yager (1998). Since the overall accuracy is just a global measure for the performance of the combination process, 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. Commission errors are the percent of incorrectly identified pixels associated with a class, and omission errors are the percent of unrecognized pixels that should have identified as belonging to a particular class. All the methods proposed in this research were implemented in Matlab (R2008b) environment.

5. RESULTS AND ANALYSIS

5.1 Comparison with Existing Fusion Algorithms

The overall classification accuracies of individual classifiers, based on the reference data, are given in Table 4. SVMs perform the best with 96.8% average overall classification accuracy, followed by SOM and CTs with average overall classification accuracies of 95.5% and 93.7% respectively. The overall accuracy of the best classifier served as a reference in the following.

Test area	Classification accuracy (%)		
	SOM	CT	SVMs
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

Table 4. Performance evaluation of single classifiers for the four test areas.

The improvement in overall classification accuracies achieved by the combination method compared with the best individual classifier, SVMs, was determined as shown in Figure 3. For the

four test areas, it is clear that the overall performances of FMV are better than those of the other combination methods. FMV performs better than WS, and both outperform MR. It is worth mentioning that even though the MR resulted in the worst performance, it still performed better than the best single classifier. Taking into account, the limited room for improvement beyond 96.9% accuracy due to other errors in image acquisition and image to lidar geographic registration, the best average improvement in classification accuracy of 1.1% is obtained from FMV algorithm, followed by 0.82% average improvement from WS algorithm. MR resulted in the worst performance and only improved the results by 0.66%. The question still remains as to whether these improvements are statistically significant. In order 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 4. With only four test areas the estimate of the SD is limited. However, 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. In the case of SOM and CTs, the higher SD values, 1.04% and 2.40% respectively, indicate that the accuracies are spread over a larger range of values for the four test areas. The SD was then used as a confidence measure in the conclusions on the quality of the accuracies derived by the three classifiers and the combined classifiers. We can assume that the reported margin of error (MOE) is typically about plus/minus twice the standard deviation (a range for an approximately 95% confidence interval). For this work we used a margin of accuracy of 0.72%, which is three times the standard deviation of the SVM results, to define the improvements in accuracy that are considered statistically significant, as shown by the dashed line in figure 3. Any improvements in classification accuracy more than the dashed horizontal line are deemed to be significant. It can be concluded that the application of FMV 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.

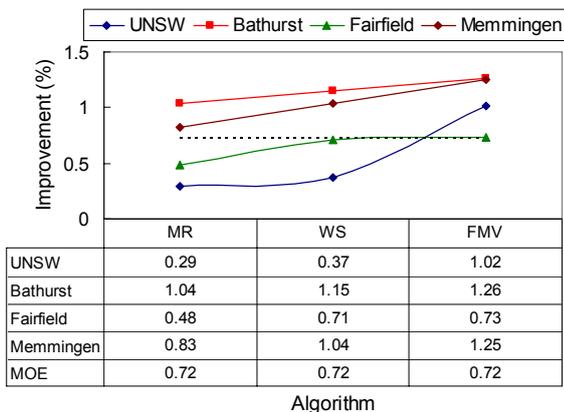


Figure 3. Performance comparison of the FMV based combination with existing algorithms, compared with the performance of the best individual classifier, SVMs. Improvements exceeding the dashed horizontal line are considered to be significant.

5.2 Class-Specific Accuracies

An assessment of the produced commission and omission errors confirms that the FMV fusion performed the best in most cases as shown in table 5. Most of the class-commission and omission errors are reduced by the FMV fusion. Whereas the application of SVMs resulted in average of 4.45 % and 5.13 % for commission and omission errors respectively, the application of FMV fusion resulted in average of 3.39 % and 2.15 % for commission and omission errors respectively. Contrary, there was an increase in commission and/or omission errors for a few classes as shown in the shaded cells of table 5. However, those classes are still classified with relatively low commission and omission errors. Another advantage of the FMV fusion over SVMs is that the achieved errors are less variable as shown in table 5. Whereas the application of SVMs resulted in standard deviation of 3.22 % and 5.25 % for commission and omission errors respectively, the application of FMV fusion resulted in a comparable SD for commission errors, 4.43%, and significantly reduced the SD for omission errors to 1.88 %. The visual assessment interpretation (Figure 4) 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 FMV appears more homogenous.

		Best Classifier		FMV Fusion	
		Com. (%)	Om. (%)	Com. (%)	Om. (%)
UNSW	B	4.65	2.77	1.31	0.82
	T	3.18	1.97	1.36	2.87
	R	4.81	0.06	0.02	3.99
	G	0.06	5.10	6.17	0.03
Bathurst	B	9.79	7.80	16.72	0.36
	T	0.35	6.12	0.02	3.82
	R	4.36	0.98	1.15	1.69
	G	10.30	4.06	9.34	1.01
Fairfield	B	8.23	11.11	3.35	1.37
	T	0.89	3.36	2.04	4.96
	R	4.08	0.76	3.41	0.01
	G	3.69	7.04	0.01	3.67
Mim	B	4.04	21.28	2.56	0.56
	T	0.63	3.94	0.85	5.36
	R	4.10	0.42	0.05	3.78
	G	7.96	5.30	5.87	0.06
Mean		4.45	5.13	3.39	2.15
SD		3.22	5.25	4.43	1.88

Table 5. Comparison of errors using the best classifier, SVMs, with the classification resulting from FMV, for the four test areas. B, T, R and G refer to buildings, trees, roads and grass respectively. Com. and Om. Refer to commission and omission errors respectively.

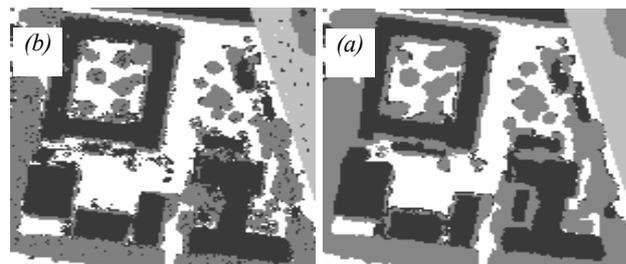


Figure 4. (a) Classification results of the best classifier (SVMs); (b) Error correction after applying the FMV fusion algorithm. Black: buildings; dark grey: trees; light grey: roads; white: ground.

6. CONCLUSION

In this paper, we have applied a powerful MCS to combine statistical and neural classifiers based on the FMV. To test the algorithm, three different classifiers based on four datasets of different sensor and scene characteristics were applied. The results showed an improvement in terms of overall classification accuracy and omission and commission errors of individual classes. Average overall accuracies of individual algorithms were 96.75%, 95.9% and 93.7% for SVMs, SOM and CTs respectively whereas the proposed fusion algorithm gives an accuracy of 97.85% which is an improvement of around 1.1%. This is an enhancement considering the limited room for improvement beyond 96.9% accuracy achieved with the SVMs, and that the data are most likely subject to other errors in image acquisition and image to lidar geographic registration, as well as errors in filtering of lidar point clouds. On the other hand, the average commission and omission errors have been reduced compared to the best single classifier. A comparison of the results with some of the existing fusion rules such as Maximum Rule (MR) and Weighted Sum (WS), demonstrates that the proposed fusion algorithm gives the best results. The computational cost involved in implementing the combined classifiers based on the FMV method is much higher than that of MR and WS methods. However, the processing time could be reduced by splitting large test areas into smaller parts, processing each part separately and combining the results later. For example, dividing the Fairfield test area, which is 4km² in area, into four equal parts can save more than 85 % of processing time (from 390 s to 58 s). The results in this paper demonstrate the overall advantages of the proposed fusion algorithm for combining multiple classifiers.

REFERENCES

- Alimohammadi, A., H.R. Rabiei, and P.Z. Firouzabadi, 2004. A new approach for modeling uncertainty in remote sensing change detection process, *Proceedings of the 12th International Conference on Geoinformatics - Geospatial Information Research: Bridging the Pacific and Atlantic*, 07-09 June 2004, University of Gavle, Sweden, pp. 503-508.
- Benediktsson, J.A., Chanussot, J. and Fauvel, M., 2007. Multiple Classifier Systems in Remote Sensing: From Basics to Recent Developments: MCS 2007, LNCS 4472, (M. Haindl, J. Kittler, and F. Roli, editors), Springer Verlag, Berlin 2007, pp. 501-512.
- Breiman, L., Friedman, J., Olshen, R., 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.
- Giacinto, G., and F. Roli, 2001. Design of effective neural network ensembles for image classification purposes. *Image Vision and Computing Journal*, 19(9-10):669-707.
- Hao, H., C.L. Liu, and H. Sako, 2003. Comparison of Genetic Algorithm and sequential search methods for classifier subset selection. *Proceeding of the seventh International Conference on document Analysis and Recognition (ICDAR)*, 3-6 August 2003, Edinburgh, Scotland, pp. 765-769.
- Herrera, F. and Verdegay, J. L., 1996. A linguistic decision process in group decision making. *Group Decision Negotiation*, 5, pp. 165-176.
- Hugo, C., Capao, L., Fernando, B. and Mario, C., 2006. Meris Based Land Cover Classification with Self-Organizing Maps: preliminary results, *Proceedings of the 2nd EARSeL SIG Workshop on Land Use & Land Cover*, 28 - 30 September 2006, Bonn, Germany, unpaginated CD-ROM.
- Kanellopoulos, I., G. Wilkinson, F. Roli, and J. Austin (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.
- Le, A.C., V.N. Huynh, A. Shimazu, and Y. Nakamori, 2007. Combining classifiers for word sense disambiguation based on Dempster-Shafer theory and OWA operators. *Data & Knowledge Engineering*, 63 (2):381-396.
- Mountrakis, G., R. Watts, L. Luo, and J. Wang, 2009. Developing Collaborative Classifiers using an Expert-based Model. *Photogrammetric Engineering and Remote Sensing*, 75(7):831-844.
- 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, pp. 15-34.
- 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, pp. 183-190.
- Yu-Chang, T. and C. Kun-Shan, 2009. An adaptive thresholding multiple classifiers system for remote sensing image classification. *Photogrammetry Engineering and Remote Sensing*, 75(6):679-687.
- Zadeh, L. A., 1983. A computational approach to fuzzy quantifiers in natural languages. *Computers and Mathematics with Applications*, 9, pp.149-184.

ACKNOWLEDGEMENTS

The authors wish to acknowledge AAMHatch for the provision of the UNSW and Fairfield datasets, the Land Property Management Authority (LPMA), NSW, Australia for Bathurst data sets and the Toposys GmbH, Germany for the Memmingen datasets.