COMBINING CLASSIFIERS FOR ROBUST HYPERSPECTRAL MAPPING USING HIERARCHICAL TREES AND CLASS MEMBERSHIPS

Karoly Livius Bakos°, Paolo Gamba*, Bogdan Zagajewski°*

*Department of Electronics, University of Pavia, Pavia, Italy
° Faculty of Geography and Regional Studies, University of Warsaw, Warsaw, Poland

I. INTRODUCTION

Remote sensing data interpretation techniques presently play a key role in many earth science, environmental and conservation applications. The availability of these datasets simplifies and speeds up the procedure of carrying out several tasks in those fields. Hyperspectral images are being increasingly made available in the last few years and are slowly being used in several applications. These images are characterized by their huge feature size with data recorded at a very fine spectral resolution in tens to hundreds of narrow frequency bands. These bands provide a wealth of information regarding the physical nature of different objects in the scene imaged by the sensors. However, the high dimensionality of the data also makes it more difficult to use the data efficiently for classification.

In this paper we focus on mapping using hyperspectral data. Using multispectral imagery, the information that can be derived from the data is limited, and apart from some broad approximation of a few physical properties of the observed surfaces, it is mostly limited to the identification of the general land cover types in the imagery. With hyperspectral images, it might also be possible to distinguish between sub-classes within the general land cover types. The inconsistent spectral signature of vegetative surfaces makes however most of the existing interpretation methodologies very scene-specific. For this reason, we suggest that only adaptive techniques can be applied in order to build a generally applicable processing chain for detailed mapping. Adaptiveness is in this case applied after classification, in order to combine results by various classifiers, whose pros and cons have been widely studied in technical literature, but without a clear guidance on where use one or the other one of them for specific applications.

For supervised classification, different approaches and algorithms are available [2]. In many cases, as was shown in previous studies [3], single stage classification systems are not flexible enough to adapt to the complexity of hyperspectral datasets. Therefore, accurate mapping becomes a particularly difficult task to carry out. Although there are data dimensionality reduction techniques and classification algorithms that are reported to be superior than others in very specific test cases, other studies also shows that the performance is also dependent on the class that is being detected [4]. The methodologies that will be introduced in this paper to combine the results of multiple classifiers will address this situation by taking into account that, even on a single scene, there can be land cover classes for which different processing chains distinguish in very different ways. A simple example is when a data set with excellent mapping performances in terms of overall accuracy can be achieved for instance using support vector machines (SVM) [5] selecting as input the first 20 components of the principal component analysis (PCA)[6]. It ma be possible, however, that for the same data sets two land cover classes show higher separability when the first 15 components of the minimum noise fraction (MNF) [7] transformation image are used for classification using a simple Maximum Likelihood (ML) [8] decision rule. In case like this one, if such multiple results can be combined to make a better decision, then a more accurate classification map can be produced.

II. DECISION FUSION USING THE HIERARCHICAL TREE AND CLASS MEMBERSHIP VALUES

Hierarchical classification and ensembles are well known in data classification methods [2]. In [9], some of the authors of this paper introduced a methodology to create an ensemble of different classifiers using a hierarchical tree structure by means of a simple learning algorithm. The learning is based on the initial analysis of the available data and it optimizes the structure of a binary decision tree (BDTC) like ensemble in terms of nodes, inputs, and decision rules to be applied at each node. This can be useful when sets of data dimensionality reduction techniques and classification algorithms are already available for the user. The aim is to combine the classification results od different processing chains using an ensemble that enables to achieve higher mapping accuracy level than any of the individual processing chains.

This methodology can be improved using a simple learning mechanism that uses class membership values provided by the various classifiers. For most of the classifiers, in fact, it is possible to define a membership value which evaluates the likelihood that the pixel belongs to a specific class. The idea is to use these class membership values, each one obtained by "weak" classifiers, to create an ensemble classifier. One disadvantage of the methodology is that the different classifications must be obtained for the full scene in advance and therefore it is not computationally effective as opposed to the hierarchical tree structure estimation introduced earlier in the chapter. In order to reduce computational costs, the hierarchical approach is used for a first selection step, in order to consider only those processing chains that were identified to be suitable for the hierarchical tree structure ensemble.

The full scene is therefore classified only with the selected processing chains and class membership values for each class are stored. The learning algorithm then works as follows:

- for every pixel, the best M classifiers are identified by ranking the maxima of membership values calculated for each classifier;
- membership values are ranked and weighted using values from M to 1 respectively and mapped into a data cube;
- the class with the largest aggregated membership value assign its label to the pixel.

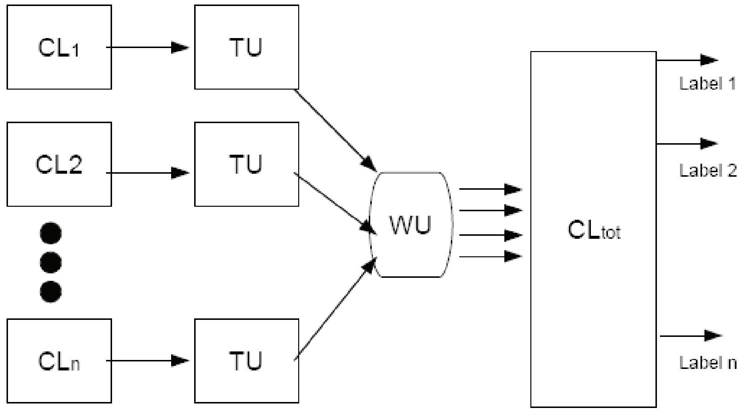

Fig. 1 The representation of class probability based ensemble classifier structure, where CL_i refers to the *i*-th classification maps, TU is a "Threshold Unit", WU is a "Weighting Unit" and CLtot is the final rule-basedclassifier.

The process is shown in Fig. 1. A so called "Threshold Unit" (TU) adaptively thresholds the soft classification results and keeps only the *M* highest probability values. Then another element, simplified as "Weighting Unit" (WU) ranks and weights the element of each of this vector of M elements with a weight from *M* to *1* and maps the values into a data cube. The data cube aggregates the weighted probability values for each class (please note that the applied weights are empirical and are used only to emphasize the most appropriate classifications). Eventually, the map is generated by rule classification of the mapped data cube by selecting the highest possible aggregated weighted probability value and labeling the pixel under test accordingly.

III. EXPERIMENTAL RESULTS

The results of the proposed approach for the well-known Indian Pine AVIRIS data set are shown in Fig. 2. Results are obtained considering a 9 classes mapping problem, and provide an overall accuracy of 90.90% for the new method, to be compared with the one by the best performing single stage classifier based processing chain (Maximum Likelihood after MNF rotation), 87.66% Fig. 2 allows also a visual comparison between the two corresponding maps while accuracy levels on a per class basis are provided in Table 1.

Probability label **ML after MNF**

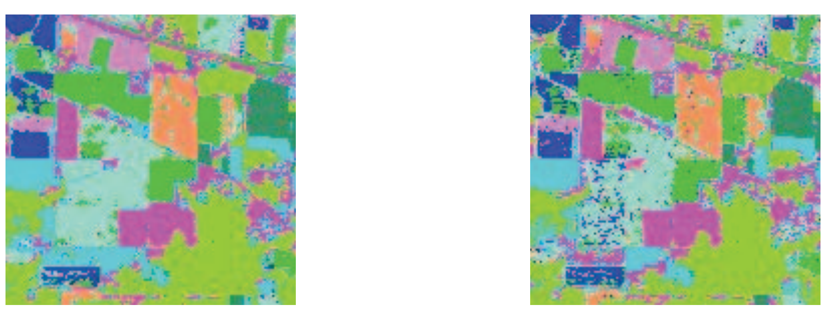

Fig. 2: The classification map obtained by using the class probability label approach compared with the

best performing single classifier.

The visual inspection of the classification maps shows that the ensemble classifier using probability labels results in a smoother classification image, containing less individual pixel errors compared to the single stage classification approach. Moreover, the field boundaries are more recognisable and less affected by pixel noise.

Table 1 The accuracy levels per class of hierarchical ensemble using probability labels against the standard hierarchical classifier in [9] and the best single stage classifier (ML after MNF rotation).

	NEW APPROACH		*HIERARCHICAL TREE*		*ML AFTER MNF*	
Class	*Prod. Acc.*	*User Acc.*	*Prod. Acc.*	*User Acc.*	*Prod. Acc.*	*User Acc.*
1	90.38	84.98	92.19	82.73	88.15	82.51
2	82.35	94.88	80.1	95.43	83.33	72.7
3	96.98	92.15	98.19	92.6	97.18	93.79
4	99.06	96.99	99.2	98.15	98.8	97.36
5	99.59	99.39	99.59	100	99.59	100
6	83.97	83.03	81.1	86.93	85.54	80
7	88.61	87.41	85.78	86.66	76.18	87.16
8	83.63	94.63	91.04	87.76	86.48	86.76
9	99.61	99.46	99.61	99.54	99.38	99.46
OA	**90.8964%**		**90.4839%**		**87.6619%**	

The new methodology provided the highest overall accuracy level among the different classifiers tested on the image. The classification image looks more realistic containing less individual pixel errors. However, as most of the classifiers are producing relatively high accuracy values, the improvement is not very high in terms of increase in the overall accuracy value.

III. REFERENCES

[1] Gamba P.,. Plaza A.J,. Benediktsson A.J, Chanussot, J., “European perspectives in hyperspectral data analysis”, Proc. of 2007 IEEE Geoscience and Remote Sensing Symposium, IGARSS 2007, Barcelona, 23-28 July 2007, pp. 4794-4797.

[2] Richards, J. A. , “Analysis of remotely sensed data: the formative decades and the future”, IEEE Trans. Geosci. Remote Sensing, vol.43, no.3, pp.422-432, 2005.

[3] Bakos, K.L., Gamba, P. “Potential of hyperspectral remote sensing for vegetation mapping in high mountain ecosystems”, Proc. of 6th EarSEL SIG IS workshop 2009, Tel Aviv, Israel, unformatted CD-ROM.

[4] L. Bruzzone, D.F. Prieto and S.B. Serpico , A neural statistical approach to multitemporal and multisensor and multisource remote sensing image classification. IEEE Trans. Geosci. Remote Sensing 37 3 (1999), pp. 1350–1359.

[5] Vladimir Vapnik, Akshay Vashist A new learning paradigm: Learning using privileged information Neural Networks, Volume 22, Issues 5-6, July-August 2009, Pages 544-557.

[6] Pearson, K. (1901). "On Lines and Planes of Closest Fit to Systems of Points in Space" (PDF). Philosophical Magazine 2 (6): 559–572.

[7] Green, A. A., Berman, M., Switzer, P., and Craig, M. D., 1988, A transformation for ordering multispectral data in terms of image quality with implications for noise removal: IEEE Transactions on Geoscience and Remote Sensing, v. 26, no. 1, p. 65-74.

[8] Richards, J.A., 1999, Remote Sensing Digital Image Analysis, Springer-Verlag, Berlin, p. 240.

[9] Bakos, K.L., Gamba, P. “A data interpretion chain for hyperspectral remote sensing data aimed at basic vegetation mapping applications”, Proc. of 2009 IEEE International Geoscience and Remote Sensing Symphosium, July 12-17, Cape Town, South Africa.