Combining Multiple Approaches for Accuracy Improvement in Remote Sensed Hyperspectral Images Classification

A. B. Santos, A. de A. Araújo
Computer Science Department
UFMG - Federal University of Minas Gerais
Belo Horizonte, MG, Brasil
Email: andreybicalho,arnaldo@dcc.ufmg.br

D. Menotti
Computing Department
UFOP - Federal University of Ouro Preto
Ouro Preto, MG, Brasil
Email: menottid@gmail.com

***Abstract*—In the past few years, hyperspectral images have been considered as one of the most important tool in land cover classification due to its capability to obtain rich information of materials on earth surface. Such amount of information can lead one to a deeply investigation, and thus a more precise interpretation. However, this task is still a challenge and producing an accurate thematic map is a common goal among researchers. In this work, we perform a combination among several classification schemes in order to improve the overall accuracy and hence build a more correct thematic map. Three types of feature representation and two learning algorithms (Support Vector Machines (SVM) and Multilayer Perceptron Neural Network (MLP)) were used yielding six classification approaches to perform the combination. Our combination proposal is based on Weighted Linear Combination (WLC), in which weights are found using a Genetic Algorithm (GA) - WLC-GA. Experiments were carried out with two well-known datasets: Indian Pines and Pavia University. They have shown that our WLC-GA achieves the highest accuracy among traditional Conscious combiners, Majority Vote (MV) and Weighted Majority Vote (WMV) for both datasets.**

***Keywords*-hyperspectral image; multiple classification systems; conscious combiners; genetic algorithm;**

I. Introduction

Remote Sensing (RS) is a field of study for gathering data and information by measuring signals from objects at long distances. In RS field, the land cover classification problem is concerned to identify different coverages on the earth surface and has been used for many purposes, such as urban planning and management, forestry, environmental monitoring, agriculture, *etc*. [1]. Hyperspectral is considered an extension of Multispectral imagery. It has spectral information expressed on hundreds of narrow bands, or channels, where each of these channels are related to a specific wavelenght. Hence, for each pixel in the image, we have a very detailed "signature" of the object [2]. Due to its rich information, hyperspectral data have shown suitable for the land cover classfication task [1]. However, with the scarce referenced data and the large amount of information, the classification task becomes more complex [3]. This scenario brings some issues, one of them is known as "curse of dimensionality" or Hughes effect [3]. Another one refers to density estimation. It is even more difficult to build this estimation when data are in high-dimensional space because this *hyperspace* is mostly empty [3]. In order to surmount such difficulties, it is very usual to apply some feature extraction/selection/representation techniques [4], [5], [6]. Among several works, most of them emphasize the fusion of both spectral and spatial information in their classification systems. *Extended Morphological Profiles* (EMP) is a simple and effective approach to encode spectral-spatial information [7], [8], [9]. The use of segmentation algorithms have also been successfully applied to this task, such as *watershed* and *Minimum Spanning Forest* (MSF) [10], [11]. Newer works use *Markov Random Fields* (MRF) for spectra-spatial classification [12], [13] and have demonstrated promising results in terms of accuracy.

Recent advances in RS field have shown *kernel*-based learning algorithms more effective than traditional classification algorithms [14], [3], [15], [16]. In particular, *Support Vector Machines* (SVM) has shown suitable for hyperspectral image classification task [14]. This classifier algorithm has been pointed out by many researchers for dealing well with high-dimensional data and small training sets [14] and its effectiveness is considered as the baseline accuracy in many works [4], [2], [12], [11].

Since 2006, an important event, called Data Fusion Contest (DFC), has been organized aiming to provide data and results as reference for the entire remote sensing community [17]. In DFC'2008, the problem tackled was information fusion using multiple classification approaches for hyperspectral data. The five best methods were combined using a majority voting (MV) rule to produce a more accurate map as possible. Despite advances, the use of more sophisticated techniques can still improve classfication results and produce more accurate thematic maps.

Inspired by this idea, this work aims to perform a more elaborated combination among different classification approaches. In the literature [18], [19], we can found several methods on combining classifiers. In this work, we investigate the advantage of using a weighted linear combination (WLC) of multiple classification approaches for the land cover classification task using remoted sensed hyperspectral images (A short version of this work have been accepted for publication [20]). We use a global optimization search technique based on Genetic Algorithms (GA) to find the weights and build the final decision. We employ three feature representation techniques (two spectral-based and one spectral-spatial) with two different learning algorithms (SVM and Multilayer Perceptron Neural Network (MLP)). Thus, we obtain six classification approaches to perform the combination. Experiments were carried on two well-known datasets: Indian Pines and Pavia University obtained by the *Airborne Visible/Infrared Imaging Spectrometer* (AVIRIS) [21] and the *Reflective Optics System Imaging Spectrometer* (ROSIS) [22] sensors, respectively. Moreover, we compare our proposed combination method (WLC-GA) with traditionals conscious combiners (Maximum, Minimum, Product and Average), Majority Vote (MV) and Weighted Majority Vote (WMV).

II. Genetic Algorithms

Genetic Algorithms (GA) is an abstraction of nature biological evolution in which the concept of "adaptation" is dressed in computer systems. Since its creation, GA has been used to solve many optimization, search, and also machine learning problems [23]. GA is capable of finding sub-optimal, or optimal, solutions in large search space with an acceptable time [23]. This evolutionary process evolves a population of "individuals" to a new one, in which its individuals represent a potential solution for a given problem. Usually, individuals are represented by strings of "genes" which are usually called as "chromosomes". Chromosomes are changed by genetic-inspired mechanisms, such as recombination, mutation and selection [23]. A *fitness* function evaluates each individual and thus provide the "quality" of a candidate solution. Hence, we must design proper individual representation and fitness function according to our problem domain. After some generations we expected to find an acceptable solution for our problem.

III. Feature Representation Schemes

In this section we briefly describe two spectral-based and one spectral-spatial feature representation techniques that are used in this work.

A. Pixelwise Representation

The pixelwise feature representation is the most simplest one. It is composed of all hyperspectral bands, *i.e.*, the raw data. Note that the so-called pixelwise type of representation does not take into account any spatial information. That is, it only uses all spectral responses/bands, or features, of the pixel. In other words, the prediction (classification) happens using all pixel features but without "looking" at its neighborhood.

B. Extended Morphological Profiles (EMP)

In order to obtain an accurate classification system, spectral and spatial information should be joined [3], [2]. For this purpose [7], the concept of morphological operations [24] is applied to build the Morphological Profiles (MP). A feature extraction (FE) technique, such as Principal Component Analysis (PCA), is applied to hyperspectral data. Then, the MPs are built applying openings and closings by reconstruction with a structuring element (SE) of increasing size. Finally, p-first Principal Components and their MPs are put together to form the Extended Morphological Profile (EMP). This technique can connect similar structures of hyperspectral data becoming its within-class representation more homogeneous [7]. Details about morphological operations can be seen in [24].

C. Proposed Feature Selection by Genetic Algorithms

In some cases, reducing data dimensionality is very important to avoid Hughes effect. Nowadays, many feature extraction/selection techniques are available in the literature [4], [5], [25]. Our work also proposes a filter-based feature selection using GA. The applied feature selection technique (Feature Extraction by Genetic Algorithm - FEGA) is based on the optimization of a clustering measure that computes the "quality" of the yielded clusters from a subset of features. Similar to [5], the chromosome is a bit string that encodes absence or presence of a feature. The fitness function is defined as the ratio between the average intra- and inter-cluster distance, *i.e.*, $F = avg(d_{intra})/avg(d_{inter})$. The "optimal" subset is that one which minimizes the F value. It is expected that this optimization produces denser clusters and more distant from each other, in the feature space. Hence, from these subset of features it is suggested to be easier to build decision boundaries for a given classifier. An extension of this feature selection approach can be seen in [26], in which a more elaborated cluster validity measure and different learning algorithms are applied.

IV. Learning Algorithms

Before starting with the two learning algorithms used in this work, let us introduce a mathematical formalism for the problem of hyperspectral image classification. Let $\delta \equiv \{1, ..., n\}$ be an integer set which indexes the n pixels of a hyperspectral image. Let $\psi \equiv \{1, ..., K\}$ be a set of K available classes and $x \equiv (x_1, ..., x_n) \in \Re^{d \times n}$ be the pixels that compose the feature vector in a d-dimensional space. Finally, let $y \equiv (y_1, ..., y_n) \in \psi^n$ represent a labeled image. The classification goal is, for every pixel $l \in \delta$, to infer a label $y_l \in \psi$ using its feature vector $x_l \in \Re^d$.

A. Multilayer Perceptron Neural Network

An *Artificial Neural Network* (ANN) of *Multilayer Perceptron* type is an extension of common ANN *Perceptron*. The MLP is composed of a set of input *units*, or *neurons*, that represents the *input layer*, at least a *hidden layer* and an *output layer* [25]. In pattern classification, a MLP separates the *feature space* using *hyperplanes*, by means of a supervised learning process. Thus, regions, in feature space, are associated with a class, then a new sample can be labeled according to the region in which it is inserted. As MLPs can have a greater number of layers, they are able to perform multiple separations in feature space. Hence, MLP can build arbitrary shapes in feature space representing different and complex classes [25]. The construction of a MLP has some issues such as the number of hidden layers and neurons in each layer, which should be set according to the problem.

We use the Neural Network Toolbox from MATLAB for the MLP classifier. In order to determine the number of neurons in the hidden layer we follow a simple rule: the square root number of input patterns times output patterns. Sigmoidal transfer function and *Backpropagation* training algorithm arethen used.

B. Support Vector Machines

The *Support Vector Machines* (SVM) methodology is based on class separation through *margins* in which samples are mapped using a *kernel* function to a higher feature space where it is expected to achieve a linear separability of data [25]. Some popular kernels are: *Linear*, *Polynomial* and *Radial Basis Function* (RBF). The ability of separating data with nonlinear distributions is related to the choice of this function and should be chosen according to the problem domain [25]. Using an appropriate nonlinear mapping, samples of two classes can then be linearly separated by a *hyperplane* [25], [27] in this new transformed and high feature space [25], [15].

Thus, SVM training consists of finding an optimal hyperplane where the separating distance between margins of each class can be maximized [25], [27]. Samples whose locations are located on the margins are called *support vectors* and are the most informative ones to build the classification decision boundary [25].

In this work, we use the LIBSVM [28] implementation for SVM classifier. RBF kernel is chosen in all feature representation schemes described in Section III and all parameters were manually adjusted based on previously experiments.

V. Combining Multiple Approaches

The main goal on combining multiple classification approaches is to produce a final decision that is better than a single one [29], [19]. Among all combiners, majority vote is the simplest one [30] and largely used [17]. However, it only uses the hard assignment which can not take advantage of all available information.

In general classifiers can produce hard or soft labels as outputs. Hard labeling uses a single *crisp* value as output, while in soft labeling for each class we have a certain degree of support [18]. In this work, we are interested on classifiers that produce soft outputs, which can provide more information that could be used for post processing. These outputs can be fuzzy, posterior probabilities, certainty, or possibility values [30], that is, for an input sample there are c membership values associated to c classes. From these soft outputs one can build a Decision Profile (DP) [18] as illustrated on Fig. 1.

Mathematically, a DP for a given sample x can be defined as a $L \times c$ matrix: $DP(x) = [D_1(x), D_2(x), ...D_L(x)]^T$ where L is the number of classifiers. Each classifier prediction is represented by a vector $D_i(x) = [d_{i,1}(x), ..., d_{i,c}(x)]$ with c classes, which $d_{i,j}(x)$ is a degree of support given a classifier D_i and a class j [18], [30]. After building support degrees

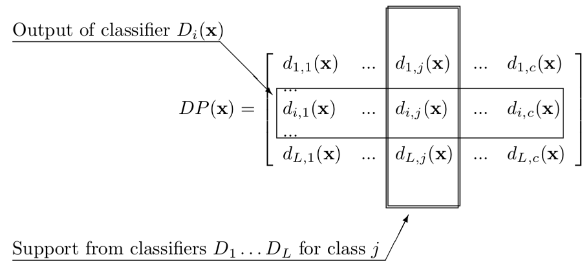


Fig. 1. Decision Profile. Extracted from [18]

for each input sample, a crisp value (hard labeling) can be assigned by using the maximum support value in the set.

In Conscious theory there are some basic rules that can be used to produce a combination between two or more decisions in DP scheme. These basic rules are defined as below:

- Maximum (Max): $\mu_j(x) = \max_i \{d_{i,j}(x)\}$
- Minimum (Min): $\mu_j(x) = \min_i \{d_{i,j}(x)\}$
- Average (Aveg): $\mu_j(x) = \frac{1}{L}\sum_{i=1}^{L} d_{i,j}(x)$
- Product (Prod): $\mu_j(x) = \prod_{i=1}^{L} d_{i,j}(x)$

where $\mu_j(x)$ is the new support degree for class j given a sample x. These operations are usually called as *nontrainable* rules just because they are conceived with no need of parameter estimation.

A. Proposed WLC-GA

When classifiers have different accuracies, it is reasonable to give more discriminant power for those that have greater accuracy [18]. This is the idea of Weighted Average, Weighted Majority Vote (WMV), and other weighted approaches [18]. Such methods are known as *trainable* because they need to find the best set of weights to produce the best set of support for each sample. In [19], the *sum* rule has proven to be a better option than other basic combination rules discussed previously. Hence, we propose to use the *sum* rule to perform a Weighted Linear Combination (WCL) in DP scheme optimized by GA (WLC-GA) [20]. Thus, new support values are produced for each class as follow:

$$\mu_j(x) = \sum_{i=1}^{L} w_i \times d_{i,j}(x)$$

in which L is the number of classifiers, w_i is the weight of the i-th classifier, $d_{i,j}(x)$ is i-th classifier support for class j and $\mu_j(x)$ is the new support for class j. Then, a label is assigned, for a given sample x, as the index of the maximum support $\mu_j(x)$. Weights are found using a global search performed by a GA. We used a bit string representation to encode the weights in the individual chromosome. Fig. 2 ilustrates our representation scheme. Each weight can be a non-negative integer value between 0 and 127, which means that there are

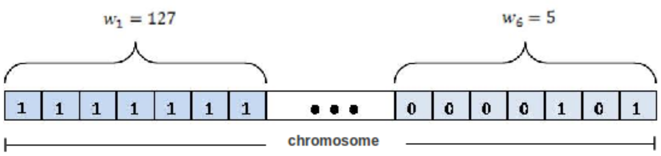


Fig. 2. Individual's representation example

7 bits in chromosome for each weight. The *fitness* function is built based on the Overall Accuracy (OA) maximization by using the WLC in the dataset.

We applied two well-known machine learning algorithms described in Section IV to each feature representation method described in Section III. Thus, as Fig. 3 shows, we obtained six classification approaches to perform the combination.

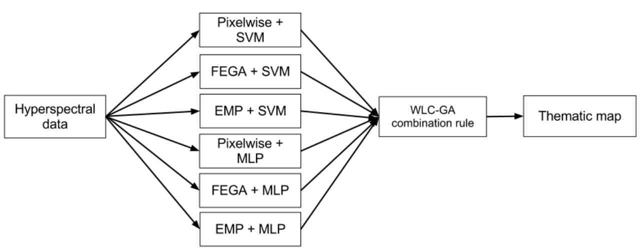


Fig. 3. Multiple classification approach scheme

VI. Experiments

A. Datasets

Experiments were performed on two widely used datasets, Indian Pines and Pavia University, taken by AVIRIS [21] and ROSIS [22] sensors, respectively. The two datasets are described below:

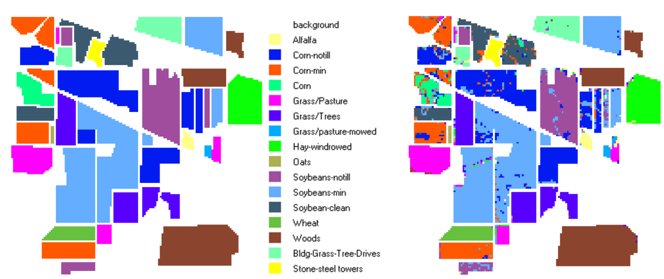
1) The Indian Pines dataset is an agricultural area recorded over Northwestern Indiana with dimensions of 145×145 pixels, a spatial resolution of $20m$ per pixel and 220 channels covering a spectral range of $0.4\mu m$ to $2.5\mu m$. Twenty noisy bands have been removed, remaining 200 spectral bands for the experiments. Sixteen classes of interest are considered, for each classification approach 5% of each class were selected as training samples. However, for those classes that have a number of samples smaller than 100 units, we selected 50% for training as shown in Table I.
2) The Pavia University dataset is an urban area recorded over the University of Pavia, Italy. The image is composed of 610×340 pixels with spatial resolution of $1.3m/pixel$, spectral range of $0.43\mu m$ to $0.86\mu m$ along its 115 channels. Therefore, 12 noisy channels were removed remaining 103 bands for the experiments. Nine classes of interest are reported, Table II summarizes all samples for each class. Again for each classification approach 5% of each class were selected as training samples.

More details about these datasets are found in [31].

To perform the combination GA's parameters were the same for both datasets used. We evolved the GA during 50 generations with a population of 50 individuals, 80% and 0.9% of *crossover* and, *mutation* probabilities, respectively, k from tournament and elitism parameters equal to 2. GA was ran in a 5-fold cross-validation scheme and the average of the weights were used as final weights to perform the combination.

B. Results

In Table I, we show the overall (OA) and average (AA) accuracy of each individual classification approach and also the proposed combination method (WLC-GA), for the Indian Pines dataset. Moreover, we compare our WLC-GA with other nontrainable combiners, such as Max, Min, Prod, Aveg, simple majority vote (MV) and we have also implemented a trainable combiner, the weighted majority vote (WMV-GA), in which the weights were also found using GA in the same scheme. The maximum reachable column (Max. reach.) represents the correct predictions union of all six approaches used, that is, if we had a combination approach that could take all true predictions of each classification method and put it together, we would get the maximum reachable accuracies for this combination method. As we expected the genetic algorithm always find higher weights for classification approaches that have highest accuracy. It is noticeable that the proposed combiner significantly improved the accuracies, however, as the maximum reachable column in Table I shows, we still can go further. The thematic map built by WLC-GA can be seen in Fig. 4b.Table II shows the results obtained for the Pavia University dataset. Again, we can observe that the proposed combiner obtained the highest accuracies. Fig. 5b shows the resulted thematic map and Fig. 5a its respective ground-truth.



(a) Ground-truth (b) Thematic map (OA-91.6%)

Fig. 4. Results for *Indian Pines* dataset

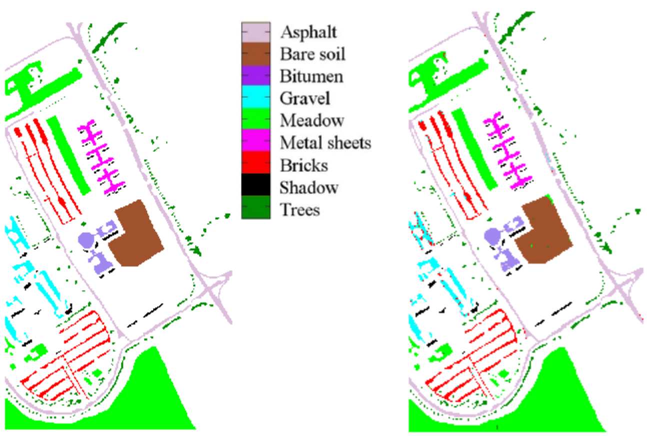
Among individual classification approaches the classification based on EMP is that one which stands the highest accuracy in both datasets. From that, we highlight the importance of having a spectral-spatial classification method in our ensemble of classifiers. Note that all accuracies are obtained from the classification maps (thematic maps).

TABLE I
RESULTS FOR INDIAN PINES DATASET

			Individual						Combination Methods							
	number of samples	train samples (%)	Pixelwise MLP (200 features)	FEGA MLP (95 features)	EMP MLP (18 features)	Pixelwise SVM (200 features)	FEGA SVM (95 features)	EMP SVM (18 features)	Max	Min	Prod	Aveg	MV	WMV -GA	WLC -GA	Max. reach.
OA (%)			79.9	77.9	81.9	81.6	77.2	88.3	89.2	82.6	87.6	89.2	87.3	90.2	91.6	97.3
AA (%)			82.3	81.3	84.6	81.8	77.0	87.8	89.6	84.6	89.6	89.3	87.1	90.6	91.6	96.8
Classes																
Alfafa	54	50	98.2	87.0	92.6	94.4	90.7	88.9	92.6	90.7	96.3	94.4	96.3	92.6	92.6	98.2
Corn-notill	1434	5	74.4	73.6	80.5	78.0	65.1	80.1	87.6	77.6	83.0	84.7	85.6	82.8	85.6	95.7
Corn-min	834	5	68.6	59.8	79.9	61.5	56.5	73.0	84.6	82.6	85.5	79.1	75.8	84.8	86.9	97.7
Corn	234	5	52.6	57.3	68.4	26.1	16.7	45.7	59.4	71.8	71.8	59.0	49.6	63.3	63.3	76.9
Grass/pasture	497	5	64.8	88.5	87.1	88.5	84.7	90.3	89.5	71.4	87.3	91.2	91.2	89.9	88.7	97.3
Grass/trees	747	5	95.4	91.8	95.7	92.6	89.2	96.0	96.8	99.0	98.6	97.1	96.8	96.8	97.5	98.2
Grass/pasture-mowed	26	50	100.0	96.2	96.2	96.2	96.2	96.2	96.1	96.1	96.1	96.1	96.1	96.1	96.1	100
Hay-windrowed	489	5	90.3	88.3	98.0	92.8	90.8	97.8	95.9	96.9	96.5	95.3	94.2	97.8	97.6	98.7
Oats	20	50	100.0	100.0	100.0	100.0	85.0	100	100	100	100	100	100	100	100	100
Soybeans-notill	968	5	68.1	75.6	66.3	75.9	71.2	84.8	78.0	71.2	71.5	84.1	85.9	85.4	85.7	94.8
Soybeans-min	2468	5	83.7	73.6	86.6	86.3	85.3	96.5	91.6	86.8	88.7	92.8	89.8	94.9	95.5	99.3
Soybean-clean	614	5	83.9	84.0	10.1	80.9	80.0	67.3	81.6	40.2	84.2	85.7	84.0	72.9	82.9	94.9
Wheat	212	5	98.1	99.1	99.1	93.4	93.9	98.1	99.5	100	99.5	99.1	98.6	98.6	99.1	99.5
Woods	1294	5	95.9	97.4	99.3	97.4	96.3	99.5	99.0	98.8	99.1	98.9	98.2	99.3	99.5	99.8
Bldg-grass-trees-drives	380	5	44.5	29.4	95.5	44.7	30.8	91.6	82.9	70.2	75.0	71.6	52.1	95.0	95.0	98.2
Stone-steel towers	95	50	98.9	98.9	98.9	100.0	100.0	98.9	98.9	100	100	100	100	98.9	100	100

TABLE II
RESULTS FOR PAVIA DATASET

			Individual						Combination Methods							
	number of samples	train samples (%)	Pixelwise MLP (103 features)	FEGA MLP (34 features)	EMP MLP (18 features)	Pixelwise SVM (103 features)	FEGA SVM (34 features)	EMP SVM (18 features)	Max	Min	Prod	Aveg	MV	WMV -GA	WLC -GA	Max. reach.
OA (%)			92.3	91.6	96.7	93.4	91.9	96.6	97.5	98.0	97.8	96.5	95.3	97.7	98.1	99.6
AA (%)			89.3	88.0	94.8	90.4	88.2	94.7	95.0	96.3	95.9	94.5	92.4	96.1	96.9	99.0
Classes																
Asphalt	6631	5	93.0	91.9	97.4	94.2	92.9	97.6	98.2	98.3	98.4	97.5	97.1	97.8	98.3	99.8
Meadow	18649	5	97.5	97.3	99.6	97.9	97.6	99.3	99.9	99.9	99.9	99.4	99.3	99.7	99.8	99.9
Gravel	2099	5	74.4	72.6	85.4	74.0	71.5	77.2	90.9	88.9	87.6	84.6	81.9	85.1	89.6	97.7
Trees	3064	5	86.7	90.1	91.6	90.2	87.0	94.4	92.3	97.6	96.8	93.0	90.8	95.7	96.2	99.2
Metal Sheets	1345	5	99.3	98.5	97.0	98.9	98.7	99.5	96.9	99.7	99.5	99.3	99.3	99.4	99.3	99.6
Bare Soil	5029	5	88.0	86.8	97.1	89.7	85.7	96.5	98.2	98.7	97.7	94.4	91.5	98.0	98.6	99.9
Bitumen	1330	5	79.5	72.9	93.5	79.8	73.8	94.4	91.0	88.5	88.4	87.8	82.0	94.1	93.3	95.9
Bricks	3682	5	85.8	82.9	91.5	88.8	87.7	93.1	95.0	94.9	95.4	94.3	90.3	95.4	96.5	99.0
Shadow	947	5	99.9	99.3	100.0	99.7	99.2	100.0	92.9	99.9	99.9	100	99.7	100	100	100



(a) Ground-truth (b) Them. map (OA-98.1%)

Fig. 5. Results for *Pavia Univeristy* dataset

C. Computation Load

We measured the computation time to perform training and testing on each of the classification methods including combination approaches. These results are shown in Tables III and IV. In combination approaches, the "training" step is concerned to find optimal, or sub-optimal, weights, while the "testing" phase performs the combination rule with the found weights. Note that simple conscious combiners do not have training phase. From these tables, we can observe that both WMV-GA and WLC-GA have higher computation load for training phase, however this time can be minimized using a different configuration of GA's parameters. On the "testing" phase, which means to apply the combination directly, the computation times for most of all combination rules in both datasets are less than 1.5 seconds, only for simple Aveg we obtained the highest computation times.

VII. CONCLUSION

In this work we proposed, implemented, and tested a combiner of classifiers for remote sensed hyperspectral images classification. We used three types of feature representation and two learning algorithms yielding six classification approaches. In the experiments, the weighted linear combiner adjusted by a genetic algorithm was also compared with the most well-known combiners in the literature and it has shown better effectiveness in terms of overall accuracy and average accuracy for the two datasets used. Although our proposed approach presents a high computation load, the use of genetic algorithm is suitable when we have a large number of classification approaches, which implies in more weights to be adjusted. In this situation, despite the fact that the search space becomes larger as the number of weights increases, the fitness function cost remains the same in our proposed scheme.

Here, we employed a simple weighted linear combination based on sum rule which uses a single weight per classifier. These results encourage us to deeply study the use of more sophisticated combiners. We plan to explore some meta-learning strategies, in which the goal is to overcome difficulties of each classification method learning how they are correlated with each other in order to improve the accuracy further. Moreover, further experiments which include mean

TABLE III
COMPUTATION LOAD (IN SECONDS) FOR TRAINING AND TESTING ON INDIAN PINES DATASET

	Pixelwise MLP	FEGA MLP	EMP MLP	Pixelwise SVM	FEGA SVM	EMP SVM	Max	Min	Prod	Aveg	MV	WMV-GA	WLC-GA
Training	63.95	60.0	34.6	0.67	0.44	0.28	-	-	-	-	-	1080	1075
Testing	0.29	0.19	0.09	2.72	1.62	0.71	0.59	0.61	0.09	3.73	0.06	0.04	0.04

TABLE IV
COMPUTATION LOAD (IN SECONDS) FOR TRAINING AND TESTING ON PAVIA UNIVERSITY DATASET

	Pixelwise MLP	FEGA MLP	EMP MLP	Pixelwise SVM	FEGA SVM	EMP SVM	Max	Min	Prod	Aveg	MV	WMV-GA	WLC-GA
Training	96.3	72.15	77.1	1.38	1.12	0.55	-	-	-	-	-	3635	3625
Testing	0.6	0.3	0.22	7.28	4.18	1.95	1.31	1.31	0.15	8.30	0.21	0.11	0.11

and standard deviation should validate more thoroughly our proposal.

ACKNOWLEDGMENT

The authors would like to thank FAPEMIG, CAPES and CNPq for the financial support.

REFERENCES

[1] C. Chang, *Hyperspectral data exploitation: theory and applications*. Wiley-Blackwell, 2007.

[2] Y. Tarabalka, J. Benediktsson, J. Chanussot, and J. Tilton, "Multiple spectral-spatial classification approach for hyperspectral data," *IEEE Trans. on Geoscience and Remote Sensing (TGARS)*, vol. 48, no. 11, pp. 4122–4132, 2010.

[3] A. Plaza, J. Benediktsson, J. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sensing Environmet*, vol. 113, no. 1, pp. 110–122, 2009.

[4] B.-C. Kuo, C.-H. Li, and J.-M. Yang, "Kernel nonparametric weighted feature extraction for hyperspectral image classification," *IEEE Trans. on Geoscience and Remote Sensing (TGARS)*, vol. 47, no. 4, pp. 1139–1155, 2009.

[5] Y. Bazi and F. Melgani, "Toward an optimal svm classification system for hyperspectral remote sensing images," *IEEE Trans. on Geoscience and Remote Sensing (TGARS)*, vol. 44, no. 11, pp. 3374–3385, 2006.

[6] L. Shen and S. Lia, "Three-dimensional gabor wavelets for pixel-based hyperspectral imagery classification," *IEEE Trans. on Geoscience and Remote Sensing (TGARS)*, vol. PP, no. 99, pp. 1–8, 2011.

[7] J. Benediktsson, J. Palmason, and J. Sveinsson, "Classification of hyperspectral data from urban areas based on extrended morphological profiles," *IEEE Trans. on Geoscience and Remote Sensing (TGARS)*, vol. 43, no. 3, pp. 480–491, 2005.

[8] J. Palmason, J. Benediktsson, J. Sveinsson, and J. Chanussot, "Classification of hyperspectral data from urban areas using morphological preprocessing and independent component analysis," in *Geoscience and Remote Sensing Symposium, 2005. IGARSS'05. Proceedings. 2005 IEEE International*, vol. 1. IEEE, 2005, pp. 4–pp.

[9] M. Fauvel, J. Chanussot, J. Benediktsson, and J. Sveinsson, "Spectral and spatial classification of hyperspectral data using svms and morphological profiles," *IEEE Trans. on Geoscience and Remote Sensing (TGARS)*, vol. 46, no. 11, pp. 3804–3814, 2008.

[10] Y. Tarabalka, J. Chanussot, and J. Benediktsson, "Segmentation and classification of hyperspectral images using watershed transformation," *Pattern Recognition*, vol. 43, no. 7, pp. 2367–2379, 2010.

[11] ——, "Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers," *IEEE Transactions on Systems, Man and Cybernetics. Part B. Cybernetics*, vol. 40, no. 5, pp. 1267–1279, 2010.

[12] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. Benediktsson, "Svm- and mrf-based method for accurate classification of hyperspectral images," *IEEE Geoscience and Remote Sensing Letters (GRSL)*, vol. 7, no. 4, pp. 736–740, 2010.

[13] B. Zhang, S. Li, X. Jia, L. Gao, and M. Peng, "Adaptative markov random field approach for classification of hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters (GRSL)*, vol. PP, no. 99, pp. 973–977, 2011.

[14] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. on Geoscience and Remote Sensing (TGARS)*, vol. 42, no. 8, pp. 1778–1790, 2004.

[15] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. on Geoscience and Remote Sensing (TGARS)*, vol. 43, no. 6, pp. 1351–1362, 2005.

[16] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. Vila-Francés, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters (GRSL)*, vol. 3, no. 1, pp. 93–97, 2006.

[17] G. Licciardi, F. Pacifici, D. Tuia, S. Prasad, T. West, F. Giacco, C. Thiel, J. Inglada, E. Chritophe, J. Chanussot, and P. Gamba, "Decision fusion for the classification of hyperspectral data: Outcome of the 2008 GRS-S data fusion contest," *IEEE Trans. on Geoscience and Remote Sensing (TGARS)*, vol. 47, no. 11, pp. 3857–3865, 2009.

[18] L. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms. 2004*. Wiley-Interscience, 2004.

[19] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 3, pp. 226–239, 1998.

[20] A. B. Santos, A. de A. Araújo, and D. Menotti, "Combining of classifiers using genetic algorithm for classification of remote sensed hyperspectral images," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2012, to appear.

[21] R. Green, M. Eastwood, C. Sarture, T. Chrien, M. Aronsson, B. Chippendale, J. Faust, B. Pavri, C. Chovit, M. Solis *et al.*, "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (aviris)," *Remote Sensing of Environment*, vol. 65, no. 3, pp. 227–248, 1998.

[22] P. Gege, D. Beran, W. Mooshuber, J. Schulz, and H. Van Der Piepen, "System analysis and performance of the new version of the imaging spectrometer rosis," in *Proceedings of the First EARSeL Workshop on Imaging Spectroscopy*. University of Zurich Remote Sensing Laboratories, 1998, pp. 29–35.

[23] D. Goldberg, *Genetic algorithms in search, optimization, and machine learning*. Addison-wesley, 1989.

[24] J. Serra, *Image Analysis and Mathematical Morphology*. London.: Academic Press., 1982.

[25] R. Duda, P. Hart, and D. Stork, *Pattern Classification and Scene Analysis*, 2nd ed. John Wiley & Sons, 1995.

[26] A. B. Santos, C. S. F. de S. Celes, A. de A. Araújo, and D. Menotti, "Feature selection for classification of remote sensed hyperspectral images: A filter approach using genetic algorithm and cluster validity," in *The 2012 International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV'12)*, 2012, to appear.

[27] C. Bishop, *Pattern recognition and machine learning*. springer New York, 2006, vol. 4.

[28] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[29] J. B. L.I. Kuncheva and R. Duin, "Decision templates for multiple classifier fusion: an experimental comparison," *Pattern Recognition*, vol. 34, pp. 299–314, 2001.

[30] A. Ghosh, B. Shankar, L. Bruzzone, and S. Meher, "Neuro-fuzzy-combiner: an effective multiple classifier system," *Int. J. of Knowledge Engineering and Soft Data Paradigms*, vol. 2, no. 2, pp. 107–129, 2010.

[31] Y. Tarabalka, "Classification of hyperspectral data using spectral-spatial approaches," Ph.D. dissertation, University of Iceland and Grenoble Institute of Technology, 2010.