A New Framework for Hyperspectral Image Classification Using Multiple Spectral and Spatial Features

*Mahdi Khodadadzadeh**, *Jun Li*†, *Antonio Plaza**, *Paolo Gamba*‡,
Jon Atli Benediktsson,§ *and José M. Bioucas-Dias*¶

*Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications,
University of Extremadura, Cáceres, E-10071, Spain.
†School of Geography and Planning, Sun Yat-Sen University, Guangzhou, P. R. China.
‡Department of Electronics, University of Pavia, Italy.
§Faculty of Electrical and Computer Engineering, University of Iceland, 101 Reykjavik, Iceland.
¶Instituto de Telecomunicações and Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal.

***Abstract*—This paper presents a new multiple feature learning approach for accurate spectral-spatial classification of hyperspectral images. The proposed method integrates multiple features based on the logarithmic opinion pool. We consider subspace multinomial logistic regression for classification as it exhibits a flexible structure for the combination of multiple features through the posterior probability. At the same time, it is able to cope with highly mixed hyperspectral data and with the presence of limited training samples. In this work, we considered lowpass filtering and morphological attribute profiles for spatial feature extraction. Our experimental results with a real hyperspectral images collected by the NASA Jet Propulsion Laboratory's Airborne Visible Infra-Red Imaging Spectrometer (AVIRIS) indicate that the proposed method exhibits state-of-the-art classification performance.**

***Index Terms*—Hyperspectral images, spectral-spatial classification, multiple features learning, subspace multinomial logistic regression (MLRsub)**

I. Introduction

In recent years, the growing availability of different remote sensing instruments has increased the possibilities of exploiting high-dimensional data in many applications [1]. Typically, for the classification of land cover classes (a main task in many applications), employing multiple features coming from different information sources has been an effecive way for remote sensing data interpretation. In other words, the explotation of data coming from multiple sources allows one to anticipate superior classification accuracies than those obtained using one source of information only. For this purpose, fusion of features coming from multiple sources has shown to be a very successful approach.

Several multiple feature learning methods haven been proposed in the literature. For example in [2], a multiple feature learning method has been suggested based on integrating information extracted from multiple spectral change difference images to build a higher quality difference image. In [3], a combined classification of a high spatial resolution color image and a lower spatial resolution hyperspectral image of the same scene has been suggested. In [4], additional features are extracted from series of multitemporal images acquired on the same scene at different dates. Other approaches extract spatial (contextual) features e.g. texture, and shape from the original data and fuse these features with original spectral features. An extensive amount of work has also been devoted to the integration of spatial and spectral features for classification of hyperspectral data. For example in [5], the authors proposed to combine multiple processing chains in a hierarchical fusion approach by using the most suitable feature extraction/ selection and classification steps. However, stacked vector based methods [6, 7] in which feature vectors are built from the concatenation of spectral and spatial features and composite kernels [8, 9] are two successful ways of integrating spatial and spectral features.

In this paper, we propose a new multiple feature learning approach for accurate spectral-spatial classification of hyperspectral images. The proposed method integrates multiple features based on the logarithmic opinion pool. We consider the subspace multinomial logistic regression (MLRsub) clasifier [10, 11], as it exhibits a flexible structure for the combination of multiple features through the posterior probability. At the same time, it is able to cope with highly mixed hyperspectral data and using limited training samples. In this work, we considered lowpass filtering and morphological attribute profiles for spatial feature extraction. In [12], it is shown that a simple lowpass filter (which is a spatial averaging operator) can reduce noise and variation within a class and

consequently, enhance class separability. Moreover, recently, morphological attribute profiles have been successfully used to model structural features by considering several of the principal components (PCs) of the image and generating an attribute profile for each of these PCs [13]. Here, we exploit the rich information provided by attribute profiles in combination with the spectral information available from the hyperspectral data to perform accurate spectral-spatial classification.

Experimental results with a real hyperspectral image collected by the NASA Jet Propulsion Laboratory's Airborne Visible Infra-Red Imaging Spectrometer (AVIRIS) indicate that the proposed approach leads to state-of-the-art classification performance for cases with very limited number of training samples.

II. Methodological Framework

The proposed approach comprises two main steps. First, spatial feature extraction to produce the additional feature vectors and then, the spatial feature vectors and the original spectral vectors from the hyperspectral data are combined in learning the MLR regressors for classification. In the following, we present the details of each step for the proposed approach.

A. Feature Extraction

Let us denote the original hyperspectral image as $\mathbf{X} \equiv (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n)$, where $\mathbf{x}_i \in \mathbb{R}^d$, for $i = 1, 2, \ldots, n$ denotes a spectral vector, n is the number of pixels in $\mathbf{x}$ and d is the number of spectral bands. Furthermore, let $\mathcal{X} \equiv (\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2, ..., \tilde{\mathbf{X}}_q)$ denote q feature images obtained by some feature extraction methods from the original image $\mathbf{X}$, *i.e.*, $\tilde{\mathbf{X}}_m = \Psi_m(\mathbf{X})$, for $m = 1, \ldots, q$. In this paper, we propose to use two strategies for extracting spatial features to use them in combination of the original spectral features. Therefore, in this work, we consider $\tilde{\mathbf{X}}_1 = \mathbf{X}$ and $q = 3$.

The first approach is to extract information from neighborhood of each pixel using lowpass filter. This strategy can increase class separability and classification accuracies, as reported in [12]. After filtering a new image $\tilde{\mathbf{X}}_2 \equiv ((\tilde{\mathbf{x}}_1)_2, (\tilde{\mathbf{x}}_2)_2, ..., (\tilde{\mathbf{x}}_n)_2)$ will be formed as follows:

$$(\tilde{\mathbf{x}}_i)_2 = \frac{1}{|N_i|} \sum_{j \in N_i} \mathbf{x}_j, \tag{1}$$

where, N_i is a neighborhood about pixel i.

The second approach is to extract structural information using morphological operations. According to [13], using morphological operations thinning and thickening with a set of thresholds λ_1, λ_2, ..., λ_n, attribute profiles (APs) of a grayscale image f are built as follows:

$$AP(f_j(\mathbf{x}_i)) = \{\phi_n(f_j(\mathbf{x}_i)), ..., \phi_1(f_j(\mathbf{x}_i)), f_j(\mathbf{x}_i), \gamma_n(f_j(\mathbf{x}_i)), ..., \gamma_1(f_j(\mathbf{x}_i))\}, \tag{2}$$

where, ϕ and γ denote the thickening and thinning transformations, respectively, and $f_j(\mathbf{x}_i)$ denotes a feature extracted from the original pixel information $\mathbf{x}_i$. Finally, by building the AP on each of the l selected components of PCA the second image $\tilde{\mathbf{X}}_3 \equiv ((\tilde{\mathbf{x}}_1)_3, (\tilde{\mathbf{x}}_2)_3, ..., (\tilde{\mathbf{x}}_n)_3)$ can be obtained based on extended attribute profiles (EAP) of the original hyperspectral image as:

$$(\tilde{\mathbf{x}}_i)_3 = \{AP(f_1(\mathbf{x}_i)), AP(f_2(\mathbf{x}_i)), ..., AP(f_l(\mathbf{x}_i))\}, \tag{3}$$

where l is the number of retained features.

B. Classification Using Multiple Spectral and Spatial Features

Let $\mathbf{Y} \equiv (\mathbf{y}_1, \ldots, \mathbf{y}_n)$ denote an image of labels, $\mathbf{y}_i \equiv [y_i^{(1)}, y_i^{(2)}, \ldots, y_i^{(k)}]^T$, where k is the number of classes. For $c = 1, \ldots, k$, if pixel i belongs to class c, $y_i^{(c)} = 1$, otherwise, $y_i^{(c)} = 0$. With these definitions in place, let $p_m(y_i^{(c)} = 1|(\tilde{\mathbf{x}}_i)_m, \boldsymbol{\omega})$ be the posterior density of associated with feature $\tilde{\mathbf{X}}_m$, where $\boldsymbol{\omega}$ is the parameter associated with the considered classifier. According to logarithmic opinion pool (LOGP) rule [14], which is a decision fusion scheme that applied to combine information from multiple features, for any pixel $i = 1, \ldots, n$, we have:

$$p_{\text{LOGP}}(y_i^{(c)} = 1|(\tilde{\mathbf{x}}_i)_1, \ldots, (\tilde{\mathbf{x}}_i)_q, \boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_q) = \frac{\prod_{m=1}^{q} p_m(y_i^{(c)} = 1|(\tilde{\mathbf{x}}_i)_m, \boldsymbol{\omega}_m)^{\alpha_m}}{\sum_{c=1}^{k} \prod_{m=1}^{q} p_m(y_i^{(c)} = 1|(\tilde{\mathbf{x}}_i)_m, \boldsymbol{\omega}_m)^{\alpha_m}}, \tag{4}$$

where $\{\alpha_m | 0 \leq \alpha_m \leq 1, \sum_{m=1}^{q} \alpha_m = 1\}$ is a tunable parameter which controls the impact of each feature vector on the final decision probability.

In this work, we use the MLR classifier to model the posterior density $p_m(y_i^{(c)} = 1|(\tilde{\mathbf{x}}_i)_m, \boldsymbol{\omega}_m)$, which is formally given by [15] as:

$$p_m(y_i^{(c)} = 1|(\tilde{\mathbf{x}}_i)_m, \boldsymbol{\omega}_m) = \frac{\exp\left(\boldsymbol{\omega}_m^{(c)} \mathbf{h}((\tilde{\mathbf{x}}_i)_m)\right)}{\sum_{c=1}^{k} \exp\left(\boldsymbol{\omega}_m^{(c)} \mathbf{h}((\tilde{\mathbf{x}}_i)_m)\right)}, \tag{5}$$

where $\mathbf{h}((\tilde{\mathbf{x}}_i)_m) \equiv [\mathbf{h}_1((\tilde{\mathbf{x}}_i)_m), \ldots, \mathbf{h}_l((\tilde{\mathbf{x}}_i)_m)]^T$ is a vector of l fixed functions of the input data, often termed as features; $\boldsymbol{\omega}_m^{(c)}$ is the set of logistic regressors for class c, and $\boldsymbol{\omega}_m \equiv [\boldsymbol{\omega}_m^{(1)^T}, \ldots, \boldsymbol{\omega}_m^{(c-1)^T}]^T$. Recently, Li *et al.* [10] proposed to combine MLR with a subspace projection method in order to develop a so-called MLRsub classifier which is able to cope with highly mixed hyperspectral data using limited training samples. In [11], a modified version of MLRsub is proposed for handling both linear and nonlinear mixtures happening at sub-pixel level in the data, as it is also common to have nonlinear mixtures in real data set. In this work, we also take

nonlinearity into account so that the input function $\mathbf{h}(\tilde{\mathbf{x}}_i)$ in (5) is given by:

$$\mathbf{h}((\tilde{\mathbf{x}}_i)_m) = [\|(\tilde{\mathbf{x}}_i)_m\|^2, \|(\tilde{\mathbf{x}}_i)_m^{\mathrm{T}}\mathbf{U}_m^{(1)}\|^2, \ldots, \|(\tilde{\mathbf{x}}_i)_m^{\mathrm{T}}\mathbf{U}_m^{(k)}\|^2]^{\mathrm{T}}, \quad (6)$$

where $\mathbf{U}_m^{(c)} = \{(\mathbf{u}_1^{(c)})_m, \ldots, (\mathbf{u}_{r^{(c)}}^{(c)})m\}$ is a set of $r^{(c)}$-dimensional orthonormal basis vectors for the subspace associated with class c ($r^{(c)} \ll d_m$, and d_m is the dimensionality of feature $\tilde{\mathbf{X}}_m$). Under the present setup, by embedding the MLRsub model in (5) into the LOGP framework for multiple feature learning in (4), we can now obtain:

$$p_{\text{LOGP}}(y_i^{(c)} = 1|(\tilde{\mathbf{x}}_i)_1, \ldots, (\tilde{\mathbf{x}}_i)_q, \boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_q) = \frac{\exp\left(\sum_{m=1}^{q} \alpha_m \boldsymbol{\omega}_m^{(c)} \mathbf{h}((\tilde{\mathbf{x}}_i)_m)\right)}{\sum_{c=1}^{k} \exp\left(\sum_{m=1}^{q} \alpha_m \boldsymbol{\omega}_m^{(c)} \mathbf{h}((\tilde{\mathbf{x}}_i)_m)\right)}. \quad (7)$$

As shown in (7), the proposed approach is able to manage multiple features with different impact factors α in the MLR model. However, it should be noted that, in this work, three types of feature vectors assuming equal values of α are considered for evaluation: the original spectral vectors, the relaxed feature vectors obtained by lowpass filtering, and the spatial features obtained by morphological attribute profiles.

III. Experimental Results

In this section, we use the well known AVIRIS Indian Pines image to evaluate the proposed approach. Detailed information about this image can be found in [10]. For training purposes, 50 samples per class are randomly selected from the ground truth image and the remaining samples are used for validation. For small classes, we take half of the total ground truth samples for training. In spatial feature extraction step, for the lowpass filtering the window neighborhood used was defined using 8-pixel connectivity and for the construction of attribute profiles we used stacked vectors of 325 features which comprise different types of attributes, as described in [13]. It should also be noted that, for all classifiers, we optimized the necessary parameters and all the experiments are repeated 30 times. The average classification accuracies are reported.

Table I shows the classification results obtained by the proposed approach. To have a complete comparison, the classification results are reported for the three kinds of feature sets using the support vector machine (SVM) and MLRsub methods. As can be seen, using spatial features, the MLRsub classifier gives higher classification accuracies than SVM. However, it is noticeable that the spectral-spatial classification using multiple features improves MLRsub classifier performance. For instance, compared with the obtained results achieved by MLRsub classifier, the proposed approach obtained an overall accuracy (OA) of 92.46% which contrasts with the OA of 72.94% using the original spectral features, OA of 87.57% using features after lowpass filtering and OA of 90.57% using attribute profiles. For illustrative purposes, Fig. 1 shows the obtained classification maps for the Indian Pines image.

IV. Conclusion and Remarks

In this paper, we presented a new multiple feature learning approach for accurate spectral-spatial classification of hyperspectral images based on the integration of logarithmic opinion pool (LOGP) rule and multinomial logistic regression (MLR) algorithm. For illustrating the capability of the proposed model for learning multiple features, we used lowpass filtering and morphological attribute profiles to have two sets of spatial feature vectors in combination of the original spectral features. Our experimental results with a real hyperspectral images collected by the NASA Jet Propulsion Laboratory's Airborne Visible Infra-Red Imaging Spectrometer (AVIRIS) indicated that the proposed method exhibits state-of-the-art performance.

V. ACKNOWLEDGEMENT

This work was supported by the Portuguese Science and Technology Foundation under Projects PEst-OE/EEI/LA0008/2013 and PTDC/EEI-PRO/1470/2012.

References

[1] A. Plaza, J. A. Benediktsson, J. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, J. A. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. 110–122, 2009.

[2] P. Du, S. Liu, P. Gamba, K. Tan, and J. Xia, "Fusion of difference images for change detection over urban areas," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 4, pp. 1076–1086, Aug. 2012.

[3] G. Thoonen, Z. Mahmood, S. Peeters, and P. Scheunders, "Multisource classification of color and hyperspectral images using color attribute profiles and composite decision fusion," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 510–521, Apr. 2012.

[4] S. Prasad, L. Bruce, and H. Kalluri, "A robust multi-classifier decision fusion framework for hyperspectral, multi-temporal classification," in *IEEE International Geoscience and Remote Sensing Symposium*, vol. 2, Jul. 2008, pp. 273–276.

[5] K. Bakos and P. Gamba, "Hierarchical hybrid decision tree fusion of multiple hyperspectral data processing chains," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 1, pp. 388–394, Jan. 2011.

TABLE I

OVERALL (OA) AND AVERAGE (AA) CLASSIFICATION ACCURACIES [%] FOR THE REAL DATA SETS CONSIDERED IN EXPERIMENTS. THE BEST RESULTS ARE OUTLINED IN BOLD TYPEFACE.

	Number of samples		Original Spectral Features		Lowpass Features		Attribute Profiles		Proposed Approach
	Train	Test	SVM	MLRsub	SVM	MLRsub	SVM	MLRsub	
Overall Accuracy	697	9669	76.87	72.94	86.66	87.57	85.65	90.57	**92.47**
Average Accuracy			85.17	82.52	91.55	92.44	90.27	93.81	**95.34**
Kappa			74.00	69.64	84.93	85.93	83.83	89.31	**91.45**

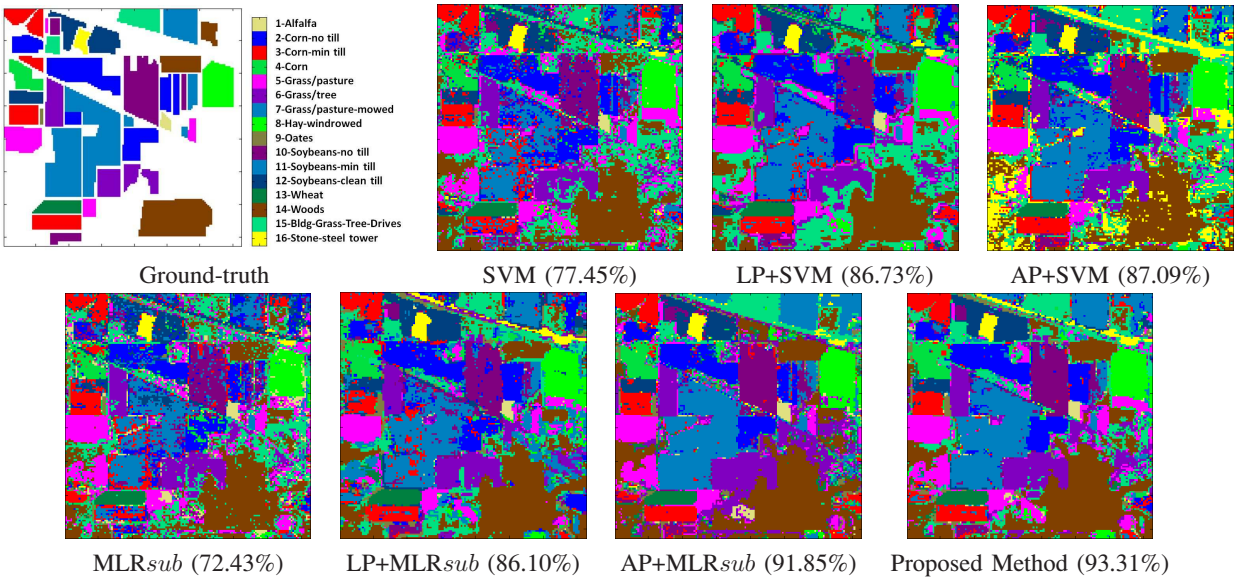

Fig. 1. Classification maps obtained by different methods for the AVIRIS Indian Pines scene (the overall accuracies are reported in the parentheses).

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

[7] P. Ghamisi, J. Benediktsson, and J. Sveinsson, "Automatic spectral-spatial classification framework based on attribute profiles and supervised feature extraction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. pp, no. 99, pp. 1–12, 2013.

[8] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. Vila-Frances, and J. CalpeMaravilla, "Composite kernels for hyperspectral image classification," *IEEE. Geoscience and Remote Sensing Letters*, vol. 3, no. 1, pp. 93–97, Jan. 2006.

[9] J. Li, P. R. Marpu, A. Plaza, J. M. Bioucas-Dias, and J. A. Benediktsson, "Generalized composite kernel framework for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 9, pp. 4816–4829, 2013.

[10] J. Li, J. Bioucas-Dias, and A. Plaza, "Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields," *IEEE Trans. Geosci. Remote Sens*, vol. 50, no. 3, pp. 809–823, 2012.

[11] M. Khodadadzadeh, J. Li, A. Plaza, and J. Bioucas-Dias, "A subspace based multinomial logistic regression for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, 2014, in press.

[12] P.-F. Hsieh and D. Landgrebe, "Classification of high dimensional data," Ph.D. dissertation, School Electr. Comput. Eng., Purdue Univ., West Lafayette, IN, 1998.

[13] M. D. Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Extended profiles with morphological attribute filters for the analysis of hyperspectral data," *International Journal of Remote Sensing*, vol. 31, no. 10, pp. 3747–3762, Oct. 2010.

[14] J. Benediktsson, J. Sveinsson, and P. Swain, "Hybrid consensus theoretic classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 4, pp. 833–843, Jul. 1997.

[15] D. Bohning, "Multinomial logistic regression algorithm," *Ann. Inst. Stat. Math.*, vol. 44, no. 1, pp. 197–200, 1992.