Semi-Supervised Classification of Urban Hyperspectral Data Using Spectral Unmixing Concepts

Inmaculada Dópido, Jun Li, Antonio Plaza
Hyperspectral Computing Laboratory
University of Extremadura
Avda. de la Universidad s/n, 10003 Cáceres, Spain
{inmaculadadopido,junli,aplaza}@unex.es

Paolo Gamba
Telecommunications and Remote Sensing Laboratory
University of Pavia
Via Ferrata 1, 27100 Pavia, Italy
Email: paolo.gamba@unipv.it

***Abstract*— Spectral unmixing and classification have been widely used in the recent literature to analyze remotely sensed hyperspectral data. However, possible connections between semi-supervised classification and spectral unmixing concepts have been rarely investigated. In this work, we propose a new method to perform semi-supervised classification of urban hyperspectral images by exploiting the information retrieved with spectral unmixing. The proposed approach integrates a well-established discriminative classifier (multinomial logistic regression) with two different spectral unmixing chains, thus bridging the gap between unmixing and classification. Moreover, the proposed method uses active learning when generating new unlabeled samples for classification.**

I. Introduction

Spectral unmixing and classification are two active areas of research in hyperspectral data interpretation. On the one hand, spectral unmixing is a fast growing area in which many algorithms have been recently developed to retrieve pure spectral components (*endmembers*) and determine their abundance fractions in mixed pixels, which dominate hyperspectral images [1]. On the other hand, supervised hyperspectral image classification is a difficult task due to the unbalance between the high dimensionality of the data and the limited availability of labeled training samples in real analysis scenarios. While the collection of labeled samples is generally difficult, expensive and time-consuming, unlabeled samples can be generated in a much easier way. This observation has fostered the idea of adopting semi-supervised learning techniques in hyperspectral image classification.

The area of semi-supervised learning has experienced a significant evolution in terms of the adopted models, which comprise complex generative models [2]–[5], self learning models [6], [7], multi-view learning models [8], [9], transductive support vector machines (SVMs) [10], [11], and graph-based methods [12]. A survey of semi-supervised learning algorithms is available in [13]. The aforementioned semi-supervised algorithms generally assume that a limited number of labeled samples are available *a priori*, and then enlarge the training set using unlabeled samples, thus allowing these approaches to address ill-posed problems. However, in order for this strategy to work, several requirements need to be met. First and foremost, the new (unlabeled) samples should be generated without significant cost/effort. Second, the number of unlabeled samples required in order for the semi-supervised classifier to perform properly should not be too high in order to avoid increasing computational complexity in the classification stage. In other words, as the number of unlabeled samples increases, it may be unbearable for the classifier to properly exploit all the available training samples due to computational issues. Further, if the unlabeled samples are not properly selected, these may confuse the classifier, thus introducing significant divergence or even reducing the classification accuracy obtained with the initial set of labeled samples. In order to address these issues, it is very important that the most highly informative unlabeled samples are identified in computationally efficient fashion.

In this work, we develop a new approach to perform semi-supervised classification of hyperspectral images by exploiting the information retrieved with spectral unmixing. Our main goal is to synergize two of the most widely used approaches to interpret hyperspectral data into a unified framework which uses active learning techniques for automatically selecting unlabeled samples in semi-supervised fashion. Specifically, we use active learning to select highly informative unlabeled training samples in order to enlarge the initial (possibly very limited) set of labeled samples and perform semi-supervised classification based on the information provided by a well-established discriminative classifier (multinomial logistic regression [14]) and different spectral unmixing chains [15], [16].

II. Proposed Approach

The proposed approach consists of three main concepts: semi-supervised learning, spectral unmixing and active learning, which are described before introducing our classifier.

A. Semi-Supervised Learning

For the semi-supervised part of our approach, we use the multinomial logistic regression (MRL) classifier [14]. In this

978-1-4799-0211-8/13/$31.00 ©2013 IEEE

paper, we use assume that the MLR is implemented with a Gaussian Radial Basis Function (RBF) kernel, which is widely used in hyperspectral image classification [17]. It is important to mention that the problem to be solved by the MLR classifier, although convex, is very difficult to compute because it involves several non-smooth terms. The sparse MLR (SMLR) algorithm presented in [18] solves this problem with $O((d(K-1))^3)$ complexity, being K the number of classes. However, most hyperspectral data sets are beyond the reach of this algorithm as their analysis becomes unbearable when the number of classes increases. In order to address this issue, we take advantage of the logistic regression via variable splitting and augmented Lagrangian (LORSAL) algorithm [19] which allows to replace a difficult non-smooth convex problem with a sequence of quadratic plus diagonal l_2-l_1 problems with practical complexity of $O(d^2(K-1))$. Compared with the figure $O((d(K-1))^3)$ of the SMLR algorithm, the complexity reduction of $d(K-1)^2$ is quite significant [19], [20].

B. Spectral Unmixing

Several spectral unmixing chains are explored in [15]. These chains are based on the well-known linear mixture model [1] and fully constrained least-squares linear spectral unmixing (FCLSU). However, the unmixing-based chains discussed in [15] do not include spatial information, which is an important source of information since hyperspectral images exhibit spatial correlation between image features. Also, the study in [15] suggested that partial unmixing using mixture-tuned matched filtering (MTMF) could be an effective solution to deal with the likely fact that not all pure spectral constituents in the scene (needed for spectral unmixing purposes) are known *a priori*, but a more exhaustive investigation of partial unmixing (in combination with spatial information) is desirable. In this work, we consider two unmixing strategies addressing the aforementioned issues as discussed in [16]:

1) *Unsupervised Clustering followed by FCLSU* (ST.1), which is intended to solve the problems highlighted by endmember extraction algorithms which are sensitive to outliers and pixels with extreme values of reflectance. By using an unsupervised clustering method such as the k-means on the original information, the endmembers extracted (from class centers) are expected to be more spatially significant. Then, FLCSU-based unmixing is conducted using the resulting endmembers.
2) *Unsupervised Clustering followed by MTMF* (ST.2), which is exactly the same as the previous strategy (ST.1) but this time MTMF-based unmixing is conducted using the resulting endmembers after k-means clustering.

C. Active Learning

We used active learning to improve the selection of unlabeled samples for semi-supervised learning. In our proposed strategy, the candidate set for the active learning process (based on the available labeled and unlabeled samples) is inferred using spatial information (specifically, by applying a first-order spatial neighborhood on available samples) so that high confidence can be expected in the class labels of the obtained candidate set. This is similar to human interaction in supervised active learning, where the class labels are known and given by an expert. In a second step, we run active learning to select the most informative samples from the candidate set. This is similar to the machine interaction level in supervised active learning, where in both cases the goal is to find the samples with higher uncertainty. Due to the fact that we use a discriminative classifier (MLR) and spectral unmixing techniques, active learning algorithms which focus on the boundaries between the classes (which are often dominated by mixed pixels) are preferred. In this way, we can combine the properties of the probabilistic MLR classifier and spectral unmixing concepts to find the most suitable (complex) unlabeled samples for improving the classification results through the selected active learning strategy. It should be noted that many active learning techniques are available in the literature [21]. In this work, we use the well-known breaking ties (BT) to evaluate the proposed approach. This algorithm finds the samples minimizing the distance between the first two most probable classes.

D. Proposed Classifier

Let $\mathbf{x} \equiv (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathbb{R}^{d \times n}$ an image of d-dimensional feature vectors, $\mathbf{y} \equiv (y_1, \ldots, y_n)$ an image of labels. With this notation in mind, the proposed method can be described as follows:

$$\widehat{p}_i(y_i = k|\mathbf{x}_i) = \alpha f_1(y_i = k|\mathbf{x}_i) + (1-\alpha) f_2(y_i = k|\mathbf{x}_i), \quad (1)$$

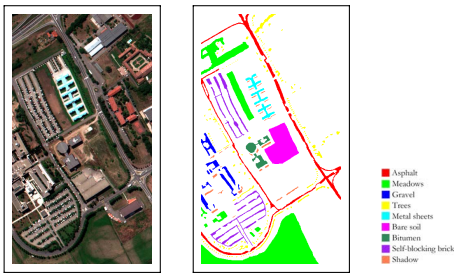
where $\widehat{p}_i(\cdot)$ is the joint estimate obtained by the classification and unmixing, $f_1(\cdot)$ is the classification probability and $f_2(\cdot)$ is the abundance fraction. The parameter α controls the degree of balance between classification and unmixing, and $0 \leq \alpha \leq 1$. If $\alpha = 1$, only classification probability is considered. If $\alpha = 0$, only unmixing has impact. Therefore, by tuning α from 0 to 1, we can adjust the impact between classification and unmixing. With the new estimate $\widehat{p}_i(\cdot)$, we expect to perform a better balanced selection of the new samples to be exploited in the semi-supervised learning chain.

III. Experimental results with Urban Hyperspectral Data Set

In this section, we use urban real hyperspectral data to evaluate the proposed approach. In all cases, the reports of overall accuracy (OA), average accuracy (AA) and κ statistic are obtained by averaging the results obtained after conducting 10 independent Monte Carlo runs on a subset randomly selected from the ground truth image, whose remaining samples are used for validation.

In order to illustrate the good performance of the proposed approach, we use on purpose very small labeled training sets. Since our approach is semi-supervised, we take advantage of unlabeled samples in order to improve classification accuracy.

All these effects are explored in the remainder of this section, organized as follows. In subsection III-A we introduce the urban dataset used for evaluation purposes in this work.



(a) (b)

Fig. 1. (a) False color composition of the ROSIS Pavia scene. (b) Ground truth-map containing 9 mutually exclusive land-cover classes.

In subsection III-B, we describe the experiments conducted with this scene. In all cases, the results obtained by the supervised version of the considered probabilistiv classifier are also reported for comparison.

A. Urban Hyperspectral data set

The urban hyperspectral data set was collected by the ROSIS optical sensor over the urban area of the University of Pavia, Italy. The flight was operated by the Deutschen Zentrum for Luftund Raumfahrt (DLR, the German Aerospace Agency) in the framework of the HySens project, managed and sponsored by the European Union. The image size in pixels is 610×340, with very high spatial resolution of 1.3 meters per pixel. The number of data channels in the acquired image is 103 (with spectral range from 0.43 to 0.86 μm). Fig. 1(a) shows a false color composite of the image, while Fig. 1(b) shows nine ground-truth classes of interest, which comprise urban features, as well as soil and vegetation features.

B. Experimental results

In this experiments, we evaluated the classification accuracy of the proposed approach using the urban ROSIS Pavia University data set. The first step was to carefully optimize parameter α controlling the impact between classification and unmixing, as we can see in the Table I. We empirically set $\alpha = 0.75$, meaning that classification should have more weight than unmixing in order to achieve higher quality results.With this parameter setting in mind, Table II shows the overall, average classification accuracies (in percentage) and the κ statistic obtained by the supervised strategy with only $\{5, 10, 15\}$ labeled samples per class and by the proposed approach (based on the MLR classifier and different spectral unmixing approaches), using two different strategies for active learning: random sampling (RS) and the BT algorithm (both executed using 300 iterations to actively select 300 unlabeled samples).

Several conclusions can be obtained from Table II. First of all, it is remarkable that the inclusion of unlabeled samples significantly improved the classification results in all cases.

For the case with 10 labeled samples per class (see Table II), the supervised approach obtained an OA of 69.25%. When the proposed strategy (combining classification and spectral unmixing) was used, the classification accuracies

TABLE I

OVERALL CLASSIFICATION ACCURACIES FOR DIFFERENT VALUES OF α WITH 5 LABELED SAMPLES PER CLASS (NUMBER OF ITERATIONS IN THE PARENTHESES).

ROSIS Pavia University					
Number of labeled samples per class ($l = 5$)					
Parameter (ite.)	1.0	0.75	0.50	0.25	0.00
ST. 1 (100)	70.17	72.03	65.00	66.85	65.32
ST. 1 (300)	75.48	79.53		70.52	
ST. 2 (100)	70.17	71.12	67.52	67.12	66.23
ST. 2 (300)	75.48	79.53	73.58		

TABLE II

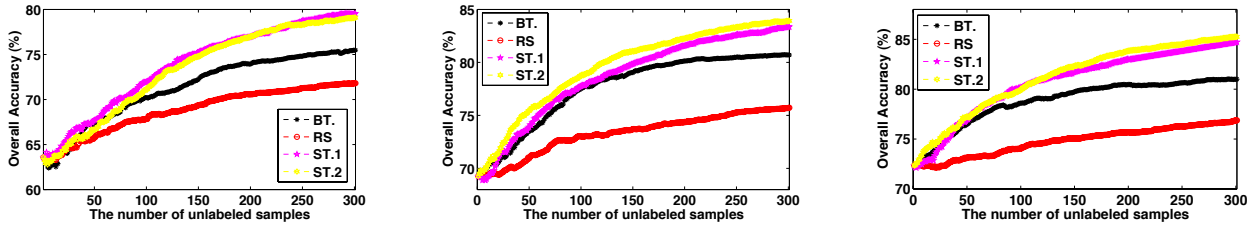
OVERALL, AVERAGE CLASSIFICATION ACCURACIES, AND κ STATISTIC BY THE MLR CLASSIFIER USING DIFFERENT APPROACHES.

	Number of labeled samples per class				
	$l = 5$				
	Supervised	BT	RS	ST.3	ST.4
OA	63.56	75.48	71.80	79.53	79.10
AA	72.93	79.33	75.45	79.97	80.29
κ	54.78	68.30	63.50	73.05	72.48
	Number of labeled samples per class				
	$l = 10$				
	Supervised	BT	RS	ST.3	ST.4
OA	69.25	80.70	75.72	83.35	83.93
AA	78.42	82.81	80.52	83.71	84.83
κ	61.69	74.87	68.95	78.12	78.97
	Number of labeled samples per class				
	$l = 15$				
	Supervised	BT	RS	ST. 1	ST. 2
OA	72.34	81.00	76.88	84.94	85.25
AA	80.01	83.35	81.61	85.59	85.48
κ	65.21	75.44	70.39	80.21	80.63

improved to an OA of 83.93% (ST.2) which represents a significant improvement over the supervised case. We also evaluated the impact of the number of unlabeled samples on the classification performance achieved. Fig. 2 shows the OAs in classification accuracy obtained by the classifier (trained using only 5, 10 and 15 labeled samples per class) and by the proposed approach (based on the classifier plus spectral unmixing) using the two considered strategies for including unmixing information and as a function of the number of unlabeled samples.

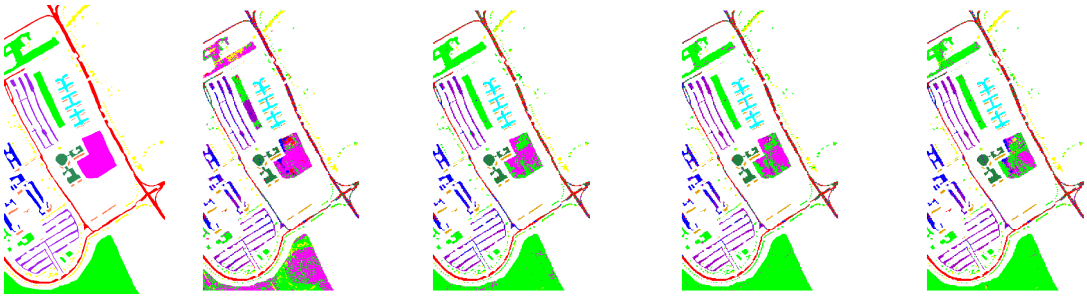
The plots in Figs. 2 reveal clear advantages of using unlabeled samples for the proposed approach when compared with the supervised algorithm alone. In all cases, the proposed unmixing strategies significantly outperform the corresponding supervised algorithm, and the increase in performance is more relevant as the number of unlabeled samples increases. In Fig. 2 it can be seen that using an intelligent active learning algorithm such BT greatly improved the obtained accuracies in comparison with simple random selection (RS).

For illustrative purposes, Fig. 3 shows some of the classification maps obtained. These classification maps correspond to one of the 10 Monte-Carlo runs that were averaged in order to generate the classification scores reported in Table II. The advantages obtained by adopting a semi-supervised learning approach which combines classification and unmixing concepts can be clearly appreciated in the classification maps displayed in Figs. 3, which also report the classification OAs obtained for each method in the parentheses.



(a) $l = 5$ (b) $l = 10$ (c) $l = 15$

Fig. 2. Overall classification accuracies as a function of the number of unlabeled samples.



Ground-truth | Supervised (63.56%) | BT with $u = 300$ (75.48%) | ST.1 with $u = 300$ (79.53) | ST.2 with $u = 300$ (79.10)

Fig. 3. Classification maps and overall classification accuracies applying the semi-supervised classifier using unmixing concepts to the ROSIS Pavia University data set (in all cases, $l = 5$).

IV. Conclusions

In this paper, a new approach, which exploits the spectral unmixing and classification concepts, has be introduced. Unlabeled samples are actively selected using active learning techniques. The effectiveness of the proposed semi-supervised approach using unmixing concepts is illustrated with real data. The results indicate that the combination of spectral unmixing and semi-supervised classification can lead to powerful new algorithms for urban hyperspectral data interpretation.

References

[1] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical and sparse regression-based approaches," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 354–379, 2012.

[2] S. Baluja, "Probabilistic modeling for face orientation discrimination: Learning from labeled and unlabeled data," in *Neural Information Procesing systems (NIPS '98)*, 1998.

[3] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training." Morgan Kaufmann Publishers, 1998, pp. 92–100.

[4] T. Mitchell, "The role of unlabeled data in supervised learning," in *Proceedings of the Sixth International Colloquium on Cognitive Science*, 1999.

[5] A. Fujino, N. Ueda, and K. Saito, "A hybrid generative/discriminative approach to semi-supervised classifer degin," in *AAAI'05 Proceedings of the 20th national coference on Artificial intelligence*, vol. 2, 2005.

[6] D. Yarowsky, "Unsupervised word sense disambiguation rivaling supervised methods," in *Proceedings of the 33rd annual meeting on Association for Computational Linguistics*, ser. ACL'95, 1995, pp. 189–196.

[7] C. Rosenberg, M. Hebert, and H. Schneiderman, "Semi-supervised self-training of object detection models," in *Seventh IEEE Workshop on Applications of Computer Vision*, January 2005.

[8] I. Cowan, T. G, and V. R. D. Sa, "Learning classification with unlabeled data," 1994.

[9] U. Brefeld, T. Gärtner, T. Scheffer, and S. Wrobel, "Efficient co-regularised least squares regression," in *Proceedings of the 23rd international conference on Machine learning*, ser. ICML'06, 2006, pp. 137–144.

[10] V. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.

[11] T. Joachims, "Transductive inference for text classification using support vector machines," in *Proceedings of the Sixteenth International Conference on Machine Learning*, ser. ICML'99, 1999, pp. 200–209.

[12] A. Blum and S. Chawla, "Learning from labeled and unlabeled data using graph mincuts," in *Proceedings of the Eighteenth International Conference on Machine Learning*, ser. ICML'01, 2001, pp. 19–26.

[13] X. Zhu, "Semi-supervised learning literature survey," Computer Sciences, University of Wisconsin-Madison, Tech. Rep. 1530, 2005.

[14] D. Böhning, "Multinomial logistic regression algorithm," *Annals of the Institute of Statistical Mathematics*, vol. 44, pp. 197–200, 1992.

[15] I. Dopido, M. Zortea, A. Villa, A. Plaza, and P. Gamba, "Unmixing prior to supervised classification of remotely sensed hyperspectral images," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, pp. 760–764, 2011.

[16] I. Dopido, A. Villa, A. Plaza, and P. Gamba, "A quantitative and comparative assessment of unmixing-based feature extraction techniques for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 421–435, 2012.

[17] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, pp. 1351–1362, 2005.

[18] B. Krishnapuram, L. Carin, M. Figueiredo, and A. Hartemink, "Sparse multinomial logistic regression: Fast algorithms and generalization bounds," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 957–968, 2005.

[19] J. Bioucas-Dias and M. Figueiredo, "Logistic regression via variable splitting and augmented Lagrangian tools," Instituto Superior Técnico, TULisbon, Tech. Rep., 2009.

[20] J. Li, J. Bioucas-Dias, and A. Plaza, "Hyperspectral image segmentation using a new Bayesian approach with active learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 19, pp. 3947–3960, 2011.

[21] D. Tuia and G. Camps-Valls, "Urban image classification with semisupervised multiscale cluster kernels," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 4, no. 1, pp. 65 –74, march 2011.