A NEW SUBSPACE DISCRIMINANT ANALYSIS APPROACH FOR SUPERVISED HYPERSPECTRAL IMAGE CLASSIFICATION

Jun Li[1,2]*, José M. Bioucas-Dias*[1] *and Antonio Plaza*[2]

[1]Instituto de Telecomunicações, Instituto Superior Técnico, TULisbon, 1900-118, Lisboa, Portugal.
[2]Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, University of Extremadura, E-10071 Caceres, Spain.

ABSTRACT

In this work, we present a new subspace discriminant analysis classification algorithm for remotely sensed hyperspectral image data. Our motivation for including subspace projection as a distinctive feature of our work is to better model noise and mixed pixels present in hyperspectral images. Two different dimensionality reduction techniques are considered: principal component analysis (PCA) and the hyperspectral signal identification by minimum error (HySime) algorithm. Experimental results indicate that the proposed method can provide competitive classification results (in the presence of very limited training data sets) with regards to those achieved by other state-of-the-art methods, such as linear discriminant analysis (LDA), subspace LDA, support vector machines (SVMs), and subspace SVMs using PCA and HySime for dimensionality reduction purposes.

Index Terms— Hyperspectral image classification, subspace analysis, sparse multinomial logistic regression, discriminant analysis.

1. INTRODUCTION

The recent availability of remotely sensed hyperspectral images in different application domains has fostered the development of techniques able to interpret this kind of high-dimensional data in many different contexts [1]. Specifically, many techniques for hyperspectral image classification have been proposed in recent years with the ultimate goal of taking advantage of the detailed information contained in hyperspectral pixel vectors (spectral signatures) to generate thematic maps [2]. A relevant challenge for the classification techniques has been the limited size of training sets versus the high dimensionality, which is known as the famous Hughes phenomenon [3]. Kernel methods, such as support vector machines (SVMs) [4], are able to circumvent this difficulty [4, 5]. However, kernel methods generally are with high computational cost.

In this work, we present a new linear subspace discriminant analysis-based approach which represents an alternative strategy for classification of remotely sensed hyperspectral data. The proposed method models the posterior class probability distributions by using a multinomial logistic regression (MLR) [6], where subspace projection is performed prior to the MLR in order to deal with the extremely high dimensionality of hyperspectral images. The regressors are learnt using a recently introduced logistic regression via splitting and augmented Lagrangian (LORSAL) algorithm [7], and the subspace projection is accomplished using two different dimensionality reduction techniques: principal component analysis (PCA) [1], and the hyperspectral signal identification by minimum error (HySime) algorithm [8]. Our experimental results indicate that the proposed method can provide competitive classification results (in the presence of very limited training data sets) with regards to those achieved by other state-of-the-art methods, such as linear discriminant analysis (LDA) [9], subspace LDA [10–12], support vector machines (SVMs) [4], and subspace SVMs using PCA and HySime for dimensionality reduction purposes.

The remaining of the paper is organized as follows. Section 2 describes the proposed subspace MLR classification method. Section 3 presents an experimental validation of the proposed subspace MLR method using two well-known hyperspectral images, collected by the Airborne Visible Infra-Red Imaging Spectrometer (AVIRIS) [13] over the Indian Pines agricultural region in NW Indiana [1], and by the Reflective Optics Spectrographic Imaging System (ROSIS) over the city of Pavia, Italy [5]. Finally, Section 4 concludes with some remarks.

2. SUBSPACE MLR CLASSIFICATION METHOD

Before describing our proposed method, we first establish the following notations:

- Let $\mathcal{S} \equiv \{1, \ldots, n\}$ denote a set of integers indexing the n pixels of a hyperspectral image.
- Let $\mathcal{L} \equiv \{1, \ldots, K\}$ be a set of K labels.
- Let $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathbb{R}^{d \times n}$ denote an image of d-dimensional feature vectors.

- Let $\mathbf{y} = (y_1, \ldots, y_n) \in \mathcal{L}^n$ be an image of labels.

With the above definitions in place, the goal of classification is to assign a label $y_i \in \mathcal{L}$ to each pixel vector $\mathbf{x}_i$ with $i \in \mathcal{S}$, resulting in an image of class labels $\mathbf{y}$. Following our previous work [14, 15], we can model the posterior densities $p(y_i|\mathbf{x}_i)$ using an MLR, which is formally given by [6]:

$$p(y_i = k|\mathbf{x}_i, \boldsymbol{\omega}) \equiv \frac{\exp(\boldsymbol{\omega}^{(k)^T}\mathbf{x}_i)}{\sum_{k=1}^{K}\exp(\boldsymbol{\omega}^{(k)^T}\mathbf{x}_i)}, \quad (1)$$

where $\boldsymbol{\omega} \equiv [\boldsymbol{\omega}^{(1)^T}, ..., \boldsymbol{\omega}^{(K)^T}]^T$ denotes the logistic regressors. Since the density in Eq. (1) does not depend on translations of the regressors $\boldsymbol{\omega}^{(K)}$, we take $\boldsymbol{\omega}^{(K)} = \mathbf{0}$ and remove it from $\boldsymbol{\omega}$, *i.e.*, $\boldsymbol{\omega} \equiv [\boldsymbol{\omega}^{(1)^T}, ..., \boldsymbol{\omega}^{(K-1)^T}]^T$. In [14, 15] we proposed different algorithms to learn the regressors by adopting different priors on $\boldsymbol{\omega}$. In this work, we introduce a subspace representation prior to the MLR in order to cope with the high dimensionality of hyperspectral images. For this purpose, we infer a p-dimensional subspace with basis $\mathbf{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$. Therefore, $\mathbf{x}$ can now be represented by $\mathbf{U}$ as $\mathbf{x} \equiv \mathbf{U}\mathbf{z}$, where $\mathbf{z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_n\}$ are coefficients. Therefore Eq. (1) turns to:

$$\begin{aligned} p(y_i = k|\mathbf{x}_i, \boldsymbol{\omega}) &\equiv \frac{\exp\left\{\boldsymbol{\omega}^{(k)^T}(\mathbf{U}\mathbf{z}_i)\right\}}{\sum_{j=1}^{K}\exp\left\{\boldsymbol{\omega}^{(j)^T}(\mathbf{U}\mathbf{z}_i)\right\}} \\ &\equiv \frac{\exp\left\{(\mathbf{U}^T\boldsymbol{\omega}^{(k)})^T\mathbf{z}_i\right\}}{\sum_{j=1}^{K}\exp\left\{(\mathbf{U}^T\boldsymbol{\omega}^{(j)})^T\mathbf{z}_i\right\}} \\ &\equiv \frac{\exp(\boldsymbol{\nu}^{(k)^T}\mathbf{z}_i)}{\sum_{j=1}^{K}\exp(\boldsymbol{\nu}^{(j)^T}\mathbf{z}_i)}, \end{aligned} \quad (2)$$

where $\boldsymbol{\nu}^{(k)} \equiv \mathbf{U}^T\boldsymbol{\omega}^{(k)}$ and $\mathbf{z}_i \equiv \mathbf{U}^T\mathbf{x}_i$. Notice that Eq. (2) is still an MLR. By adopting an l_1 prior on $\boldsymbol{\nu}$ so that $p(\boldsymbol{\nu}) \propto \exp(-\lambda\|\boldsymbol{\nu}\|_1)$, where λ acts as a tunable parameter controlling the degree of sparseness, now it is possible to infer the regressors by using the LORSAL algorithm [7].

In the following, we will refer to the standard (linear) MLR strategy in Eq. (1) as "linear MLR", and to our proposed subspace-based modification in Eq. (2) as "subspace MLR". There are several advantages of our proposed subspace MLR with regards to the linear MLR. First and foremost, the subspace MLR involves less parameters in the model and is less sensitive to the curse of dimensionality which affects hyperspectral data interpretation [1]. This allows the proposed subspace MLR to perform well in the presence of training sets made up of a very limited number of available samples. Furthermore, as will be shown in experiments the proposed subspace projection method can also reduce noise and, subsequently, it can reduce the impact of mixed pixels in the classification process. This is because noise can lead to confusion between spectrally similar classes resulting from a predominance of mixed pixels, as it is indeed the case in the hyperspectral scenes that will be used in the following section.

3. EXPERIMENTAL RESULTS

In this section, two real hyperspectral data sets are used to evaluate the proposed method in comparison with the following methods: SVM, LDA and linear MLR (using all input bands for classification purposes) versus subspace SVM, LDA and MLR (using both PCA and HySime for dimensionality reduction purposes). In all subspace-based methods, the number of subspace components was set to retain 99.99% of the variance in the original data. In the SVM experiments, only results for the Gaussian radial basis function (RBF) kernel [4] (*i.e.*, the one that resulted in the best classification results) are reported. Each value of accuracy reported in experiments is obtained from 10 Monte Carlo runs.

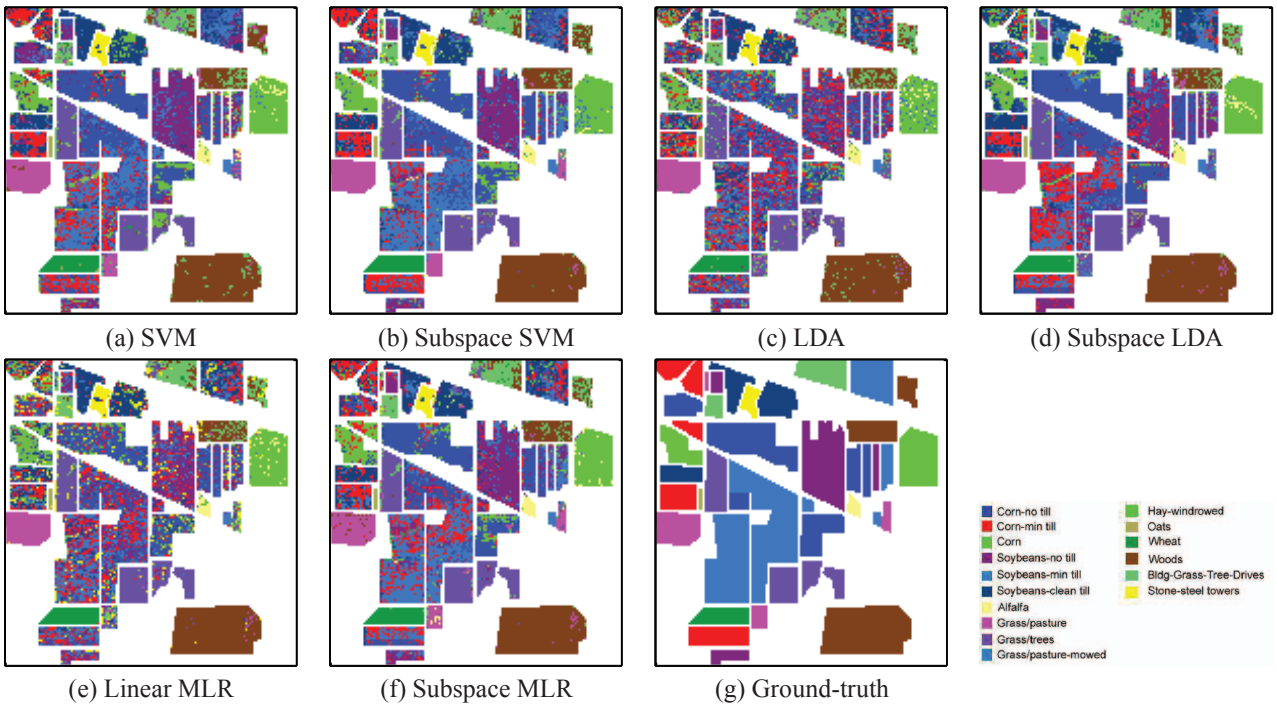
3.1. Experiments with the AVIRIS Indian Pines scene

The well-known AVIRIS Indian Pines scene was collected over Northwestern Indiana in June of 1992 [1]. It contains 145×145 pixels and 220 spectral bands, where a total of 20 bands were removed prior to experiments due to noise and water absorption in those channels. The ground-truth data contains 16 mutually exclusive classes, and a total of 10366 labeled pixels. This image is a classical benchmark to validate the accuracy of hyperspectral image analysis algorithms and constitutes a challenging problem due to the significant presence of mixed pixels in all available classes, and also because of the unbalanced number of labeled pixels per class.

Table 1 shows the overall accuracy (OA) [%], average accuracy (AA) [%], kappa statistic (κ) results and computational time (seconds) as a function of the number of labeled samples. In our experiments, we have considered different numbers of training samples $\{160,\ 240,\ 320\}$ evenly distributed among the 16 available classes (for instance, for the 320 case, 20 training samples were used per class for training purposes and the remaining samples were used for testing purposes). From Table 1, it can be seen that the subspace MLR (with HySime) obtained better results than those reported for the other tested methods in linear cases, and comparable with those obtained by the SVM-based algorithm (regardless of the number of training samples). Furthermore, it can be observed that subspace projection increased classification performance of linear approaches: LDA and MLR. Finally, HySime improved all classifiers as it removes noise and alleviates the negative impact of mixed pixels. For illustrative purposes, Fig. 1 shows a random classification map (out of the 10 Monte Carlo experiments) for each method, trained with 320 training samples, along with the ground-truth map. In all cases, the best dimensionality reduction method (HySime) is used for the subspace-based methods. Improvements in classification accuracy can be visually appreciated when comparing the subspace MLR with regards to other tested methods.

Table 1. OA [%]/AA [%]/κ and time (seconds) results as a function of the number of labeled samples for different hyperspectral image classification algorithms over 10 Monte Carlo runs in a PC with CPU at 1.60 GHz and 1 GB of RAM memory. In the case of 160 training samples, no result could be reported for LDA and linear MLR which require a number of training samples larger than the dimensionality of input feature vectors.

Classification algorithm		**Number of training samples**											
Classifier	**Dimensionality**	160				240				320			
adopted	**reduction**	OA	AA	κ	time	OA	AA	κ	time	OA	AA	κ	time
SVM	No reduction	58.14	69.78	0.53	23.54	61.00	73.04	0.56	32.09	67.86	78.07	0.64	47.48
Subspace SVM	HySime	60.69	74.44	0.56	25.58+3.40	65.56	76.77	0.61	25.58+4.78	70.17	80.25	0.67	25.58+5.59
	PCA	52.45	66.02	0.47	4.19+3.47	57.92	69.70	0.53	4.19+5.31	64.48	75.05	0.60	4.19+6.48
LDA	No reduction			-		35.77	46.37	0.29	1.09	52.54	65.10	0.47	1.39
Subspace LDA	HySime	59.82	72.45	0.55	25.58+0.28	61.53	73.97	0.57	25.58+0.30	63.00	74.90	0.59	25.58+0.31
	PCA	56.03	69.00	0.51	4.19+0.28	58.10	70.48	0.53	4.19+0.28	60.90	71.87	0.56	4.19+0.28
Linear MLR	No reduction			-		51.73	65.68	0.47	17.08	55.71	67.60	0.51	20.19
Subspace MLR	HySime	61.50	72.85	0.57	25.58+1.71	65.21	76.37	0.61	25.58+2.14	68.44	78.03	0.64	25.58+2.64
	PCA	55.62	68.17	0.51	4.19+1.70	60.39	71.30	0.56	4.19+2.08	65.12	74.53	0.61	4.19+2.64



(a) SVM (b) Subspace SVM (c) LDA (d) Subspace LDA

(e) Linear MLR (f) Subspace MLR (g) Ground-truth

Fig. 1. Classification maps obtained using $L = 320$ training samples. (a) OA = 65.36%, AA = 77.30%, $\kappa = 0.61$. (b) OA = 70.33%, AA = 81.25%, $\kappa = 0.67$. (c) OA = 50.74%, AA = 66.66%, $\kappa = 0.45$. (d) OA = 54.90%, AA = 67.27%, $\kappa = 0.50$. (e) OA = 60.38%, AA = 73.72%, $\kappa = 0.56$. (f) OA = 67.53%, AA = 79.75%, $\kappa = 0.64$. (g) Ground truth image.

Table 2. OA [%] and κ statistic (in the parentheses) for the ROSIS Pavia Centre.

Classification algorithm		Number of training samples per class						All
Classifier	**Dimensionality**	10	20	40	60	80	100	
Linear MLR	No reduction	-	79.99 (0.68)	88.62 (0.81)	90.29 (0.83)	91.05 (0.85)	91.68 (0.86)	96.90 (0.95)
Subspace MLR	HySime	89.38 (0.82)	91.80 (0.86)	94.45 (0.91)	95.80 (0.83)	96.08 (0.93)	96.34 (0.94)	97.56 (0.94)
SVM	No reduction	93.26 (0.89)	94.59 (0.91)	94.96 (0.91)	95.49 (0.92)	95.60 (0.93)	95.40 (0.92)	96.22 (0.94)
Subspace SVM	HySime	90.29 (0.94)	93.98 (0.90)	95.43 (0.92)	96.59 (0.94)	96.75 (0.94)	96.92 (0.95)	98.11 (0.97)

3.2. Experiments with the ROSIS Pavia scene

In this subsection, a ROSIS hyperspectral data set collected over the town of Pavia is used to evaluate the propose approach. Specifically, a subset with 492×1096 pixels in size, collected over Pavia city centre, is used in our experiments. Noisy bands were removed prior to the analysis, yielding a dataset with 102 spectral bands. The ground-truth image contains 9 classes, with a total of 5536 training samples and 103539 test samples. In this experiment, we only show the results obtained by the MLR and SVM classifiers, as they perform much better than the LDA classifier. For the subspace approach, we use HySime for dimensionality reduction as we experimentally observed that HySime outperformed PCA in terms of OA, κ and AA in all cases.

Table 2 shows the OA and κ results for the ROSIS Pavia scene as a function of the number of labeled samples, composed of 10, 20, 40, 60, 80 and 100 pixels per class. Notice the good performance obtained by the proposed approach. For example, with 40 labeled samples per class the proposed subspace MLR method obtained an OA of 94.45%, which is 5.83% higher than that obtained by the standard linear MLR. Furthermore, the results obtained by the proposed linear subspace MLR method are comparable with those obtained by the SVM-based algorithms however with much less computational cost. For instance, with the full training set, the proposed subspace MLR method only took 18.04 (computational time of HySime)+0.72 seconds, while the subspace SVM took 18.04+113.92 seconds and the original SVM with all spectral bands took 182.09 seconds. We can also observe that the classification results are improved for HySime subspace identification in the cases of linear and kernel-based classifiers.

4. CONCLUSIONS

In this paper, we have introduced a new subspace-based discriminant analysis approach for supervised classification of hyperspectral image data sets. In our proposed approach, the posterior class probability distributions are modeled by an MLR under subspace projection methods. We highlight the excellent classification result obtained with very limited training samples, which is competitive with regards to other similar approaches in the literature.

5. ACKNOWLEDGMENTS

This research was supported by MRTN-CT-2006-035927 and AYA2008-05965-C04-02. The authors thank D. Landgrebe and P. Gamba for providing the hyperspectral data sets.

6. REFERENCES

[1] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*, John Wiley, Hoboken, NJ, 2003.

[2] A. Plaza et al., "Recent advances in techniques for hyperspectral image processing," *Remote Sensing of Environment*, vol. 113, pp. 110–122, September 2009.

[3] G.F. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Transactions on Information Theory, IT*, vol. 14, no. 1, pp. 55–63, 1968.

[4] 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.

[5] M. Fauvel, J.A. Benediktsson, J. Chanussot, and J.R. 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, 2008.

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

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

[8] J. Bioucas-Dias and J. Nascimento, "Hyperspectral subspace identification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 8, pp. 2435–2445, 2008.

[9] Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification (2nd Edition)*, Wiley-Interscience, 2000.

[10] D. L. Swets and J. Went, "Using discriminating eigenfeatures for image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, pp. 831–836, 1996.

[11] J. Ye and Q. Li, "A two-stage linear discriminant analysis via qrdecomposition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, pp. 929–941, 2005.

[12] S. Prasad and L.M. Bruce, "Limitation of principal components analysis for hyperspectral target recognition," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, pp. 625–629, 2008.

[13] R. O. Green et al., "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sensing of Environment*, vol. 65, no. 3, pp. 227–248, 1998.

[14] J. Li, J. Bioucas-Dias, and A. Plaza, "Semi-supervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, pp. 4085–4098, 2010.

[15] J. Li, J. Bioucas-Dias, and Antonio Plaza, "Hyperspectral image segmentation using a new bayesian approach with active learning," *IEEE Transactions on Geoscience and Remote Sensing (accepted)*.