Subspace-Based Support Vector Machines for Hyperspectral Image Classification

Lianru Gao, Jun Li, *Member, IEEE*, Mahdi Khodadadzadeh, Antonio Plaza, *Senior Member, IEEE*, Bing Zhang, *Senior Member, IEEE*, Zhijian He, and Huiming Yan

***Abstract*—Hyperspectral image classification has been a very active area of research in recent years. It faces challenges related with the high dimensionality of the data and the limited availability of training samples. In order to address these issues, subspace-based approaches have been developed to reduce the dimensionality of the input space in order to better exploit the (limited) training samples available. An example of this strategy is a recently developed subspace-projection-based multinomial logistic regression technique able to characterize mixed pixels, which are also an important concern in the analysis of hyperspectral data. In this letter, we extend the subspace-projection-based concept to support vector machines (SVMs), a very popular technique for remote sensing image classification. For that purpose, we construct the SVM nonlinear functions using the subspaces associated to each class. The resulting approach, called SVMsub, is experimentally validated using a real hyperspectral data set collected using the National Aeronautics and Space Administration's Airborne Visible/Infrared Imaging Spectrometer. The obtained results indicate that the proposed algorithm exhibits good performance in the presence of very limited training samples.**

***Index Terms*—Hyperspectral image classification, multinomial logistic regression (MLR), subspace-based approaches, support vector machines (SVMs).**

I. Introduction

HYPERSPECTRAL image classification has been a very active area of research in recent years [1]. Given a set of observations (i.e., pixel vectors in a hyperspectral image), the goal of classification is to assign a unique label to each pixel vector so that it is well defined by a given class [2]. Although techniques for unsupervised classification and/or clustering have also been used in the literature [3], supervised classification has been more widely used [4], but it also faces challenges related with the high dimensionality of the data and the limited availability of training samples [2]. Even though hyperspectral images are characterized by their high spectral resolution, which allows capturing fine details of the spectral characteristics of materials in a wide range of applications [5], it has been demonstrated that the original spectral features contain high redundancy [3]. Specifically, there is a high correlation between adjacent bands, and the number of the original spectral features may be too high for classification purposes [4]. In addition, the original spectral features may not be the most effective ones to separate the objects of interest from others, since the hyperspectral data may effectively live in a lower dimensional subspace [6]. These observations have fostered the use of subspace-based techniques for hyperspectral image classification, aimed at reducing the dimensionality of the input space in order to better exploit the (often limited) training samples available *a priori*.

In the hyperspectral imaging literature, subspace-based techniques have been widely used in spectral unmixing problems [7], [8], which interpret mixed pixels in a hyperspectral scene in terms of a collection of pure spectral signatures (*endmembers* [9]) and their corresponding abundance fractions [10]. The connections between spectral unmixing and subspace projection were first explored in [11]. In [6], a technique called hyperspectral subspace identification with minimum error was presented to identify the subspace in which the hyperspectral data live, which is related with the estimation of the number of endmembers in a given scene. However, subspace-based techniques have also been used for detection and classification purposes, mainly due to their capacity to deal with mixed pixels and interferers. For instance, in [12] (later revisited in [13]), an orthogonal subspace projection technique was introduced for hyperspectral image classification and dimensionality reduction. In [14], a least squares subspace projection approach to mixed pixel classification was discussed. In [15], a kernel-based subspace projection technique was presented and evaluated in the context of an agricultural application. In [16], a technique for hyperspectral signal subspace identification in the presence of rare signal components was explored. More recently, a subspace-projection-based multinomial logistic regression (MLR) classifier, called MLRsub [17], has been presented. This classifier models the subspace associated with each specific class. In other words, the MLRsub uses a class-dependent procedure integrated with the MLR classifier to represent each class using a subspace spanned by a set of basis vectors. This approach exhibited good classification performance using several hyperspectral scenes. A general conclusion from the aforementioned studies is that subspace projection methods are useful for the separation of classes which are very similar in spectral terms due to spectral mixing and other phenomena.

Inspired by the previous development of MLRsub, this letter presents a new methodology that combines a class-indexed subspace projection technique integrated with the support vector machine (SVM) classifier [18], [19], which has been widely

Manuscript received December 16, 2013; revised April 1, 2014; accepted July 2, 2014. *(Corresponding author: Jun Li.)*

L. Gao, B. Zhang, and H. Yan are with the Key Laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China.

J. Li and Z. He are with the School of Geography and Planning and the Guangdong Key Laboratory for Urbanization and Geo-Simulation, Sun Yat-Sen University, Guangzhou 510275, China.

M. Khodadadzadeh and A. Plaza are with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, 10003 Cáceres, Spain.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2014.2341044

1545-598X © 2014 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

used in order to deal effectively with the Hughes phenomenon. This phenomenon is related with the imbalance between the high dimensionality (in spectral sense) of hyperspectral data and the (generally limited) number of training samples available, which often compromises the performance of supervised classification techniques [2]. The SVM was first investigated as a binary classifier [20]. Given a training set mapped into a space by some mapping, the SVM separates the data by an optimal hyperplane. If the data are linearly separable, we can select two hyperplanes in a way that they separate the data and there are no points between them, and then try to maximize their distance. The region bounded by them is called the margin [21]. If the data are not linearly separable, soft margin classification with slack variables can be used to allow misclassification of difficult or noisy cases. However, the most widely used approach in SVM classification is to combine soft margin classification with a kernel trick that allows separation of the classes in a higher dimensional space by means of a nonlinear transformation. In other words, the SVM used with a kernel function is a nonlinear classifier, where the nonlinear ability is included in the kernel and different kernels lead to different types of SVMs. The extension of SVMs to multiclass problems is usually done by combining several binary classifiers [20]. In this letter, our main contribution is to incorporate a subspace-projection-based approach to the classic SVM formulation, with the ultimate goal of having a more consistent estimation of the class distributions. The resulting classification technique, called SVMsub, is shown in this work to be robust to the presence of noise, mixed pixels, and limited training samples.

The remainder of this letter is organized as follows. Section II presents the proposed SVMsub classification technique. Section III evaluates the classification performance of the SVMsub method in comparison with another established subspace-based classifiers such as the MLRsub, using a hyperspectral data set collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over the Indian Pines region in Indiana. Our experimental results indicate that the proposed SVMsub algorithm provides competitive classification results in comparison with other approaches. Section IV concludes this letter with some remarks and hints at plausible future research lines.

II. Class-Dependent Subspace-Based SVM (SVMsub)

Let $\mathbf{x} \equiv \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ be a hyperspectral image, where $\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{id}]^{\mathrm{T}}$ denotes a spectral vector associated with an image pixel $i \in S$, $S = \{1, 2, \ldots, n\}$ is the set of integers indexing the n pixels of $\mathbf{x}$, and d is the number of spectral bands. Let $\mathbf{y} \equiv (y_1, \ldots, y_n)$ denote an image of class labels so that $y_i \in \{-1, 1\}$. For a label set $\mathcal{K} \equiv \{1, \ldots, K\}$, where K is the number of classes, if pixel i belongs to class k, we say $y_i^{(k)} = 1$ and $y_i^{(c)} = -1$ for $c \in \mathcal{K}$ and $c \neq k$. Using the linear mixture model assumption [8], hyperspectral data are assumed to live in class-dependent subspaces [17]. Therefore, for any $i \in S$, we can write

$$\mathbf{x}_i = \sum_{k=1}^{K} \mathbf{U}^{(k)} \mathbf{z}_i^{(k)} + \mathbf{n}_i \tag{1}$$

where $\mathbf{n}_i$ is the noise, $\mathbf{z}_i^{(k)}$ denotes the coordinates of $\mathbf{x}_i$ with respect to the basis $\mathbf{U}^{(k)}$, and $\mathbf{U}^{(k)} = \{\mathbf{u}_1^{(k)}, \ldots, \mathbf{u}_{r^{(k)}}^{(k)}\} \in \mathbb{R}^{d \times r^{(k)}}$ is a set of $r^{(k)}$-dimensional orthonormal basis vectors for the subspace, associated with classes $k = 1, 2, \ldots, K$. Under the projection principle, let $\mathbf{R}^{(k)} = E\{\mathbf{x}_{l^{(k)}}^{(k)} \mathbf{x}_{l^{(k)}}^{(k)^{\mathrm{T}}}\}$ denote the calculated matrix associated with class k, and $\mathbf{x}_{l^{(k)}}^{(k)}$ is the training set associated with class k containing $l^{(k)}$ samples; then, $\mathbf{U}^{(k)}$ is the basis representing $\mathbf{R}^{(k)}$ as $\mathbf{R}^{(k)} = \mathbf{U}^{(k)} \mathbf{\Lambda}^{(k)} \mathbf{U}^{(k)^{\mathrm{T}}}$, where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1^{(k)}, \ldots, \lambda_d^{(k)})$ is the eigenvalue matrix with decreasing magnitude, i.e., $\lambda_1^{(k)} \geq \cdots \geq \lambda_d^{(k)}$. Following [17], we determine $r^{(k)}$ so as to retain 99% of the original spectral information, i.e., $r^{(k)} = \min\{r^{(k)} : \sum_{i=1}^{r^{(k)}} \lambda_i^{(k)} \geq \sum_{i=1}^{d} \lambda_i^{(k)} \times 99\%\}$.

In [17], a nonlinear function defined as $\boldsymbol{\phi}(\mathbf{x}_i) = [\|\mathbf{x}_i\|^2, \|\mathbf{x}_i^{\mathrm{T}} \mathbf{U}^{(1)}\|^2, \ldots, \|\mathbf{x}_i^{\mathrm{T}} \mathbf{U}^{(K)}\|^2]^{\mathrm{T}}$ was used to learn the logistic regressors. With the aforementioned notation in mind, we now extend the class-dependent subspace concept to the SVM classifier and also define the implementation of the nonlinear function $\boldsymbol{\phi}(\mathbf{x}_i)$ for the SVM model. Therefore, after the nonlinear transformation, the hyperspectral data $\mathbf{x}$ turns to $\boldsymbol{\phi}(\mathbf{x}) \equiv \{\boldsymbol{\phi}(\mathbf{x}_1), \ldots, \boldsymbol{\phi}(\mathbf{x}_n)\}$ in the SVM implementation.

The SVM is a supervised nonparametric statistical learning technique which learns from a set of labeled data instances, thus trying to find an optimal hyperplane that separates the data set into a discrete (and predefined) number of classes in a way that is consistent with the training examples [19]. Here, the notion of optimal separation hyperplane refers to the decision boundary that minimizes misclassifications, which is obtained after the training step. In our context, this would be equivalent to assuming that the original spectral features are linearly separable in the input space. In practice, linear separability is difficult as the basic linear decision boundaries are often insufficient to properly model the data. For this purpose, a kernel trick is used to solve the inseparability problem by mapping the nonlinear correlations into a higher dimensional space [22].

Obviously, the choice of the input function can have a significant impact on the obtained results [4]. In our approach, by introducing the class-dependent subspace-based nonlinear function $\boldsymbol{\phi}(\mathbf{x})$, the decision rule for a pixel u can be obtained as

$$y_u = \mathrm{sgn}\left(\sum_{i=1}^{l_n} y_i \alpha_i \left(\boldsymbol{\phi}(\mathbf{x}_i)^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}_u)\right) + b\right) \tag{2}$$

where l_n is the number of labeled samples and $0 \leq \alpha_i \leq C$, with C being the soft margin parameter. For simplicity, sometimes, it is required that the hyperplane passes through the origin of the coordinate system, whereas general hyperplanes not necessarily passing through the origin can be enforced by setting $b = 0$ in (2) [21]. As shown in this equation, in our approach, the parameters involved in the calculation of the input function are K-dimensional, where K is the number of classes, independently of the size of the training set. This brings a significant advantage from a computational complexity viewpoint since, for some conventional kernel functions such as the Gaussian radial basis function (RBF) or polynomial, the size of the kernel depends on the training set l_n. This generally results in a prohibitive calculation for problems with large training sets. Therefore, the proposed approach (which adopts a linear SVM after a nonlinear transformation) significantly reduces the computational complexity and allows the SVM to manage problems with large training sets.

TABLE I
OVERALL AVERAGE INDIVIDUAL CLASS ACCURACIES (IN PERCENT) AND κ STATISTIC OBTAINED BY THE DIFFERENT TESTED METHODS FOR THE AVIRIS INDIAN PINES SCENE. IN ALL CASES, ONLY 320 TRAINING SAMPLES IN TOTAL (20 PER CLASS) WERE USED

Class	# samples	Results using 220 spectral bands			Results using 200 spectral bands		
		SVMsub	SVM	MLRsub	SVMsub	SVM	MLRsub
Alfalfa	54	85.23	75.40	84.23	84.34	86.17	83.34
Corn-no till	1434	55.78	32.35	54.34	54.42	47.22	53.25
Corn-min till	834	54.45	39.18	59.92	57.32	50.42	59.47
Corn	234	75.94	66.55	74.95	73.39	71.36	73.03
Grass/pasture	497	83.21	69.94	83.91	84.02	79.54	83.68
Grass/tree	747	90.97	71.10	91.86	90.97	80.03	90.57
Grass/pasture-mowed	26	89.87	86.87	91.53	90.20	89.53	92.20
Hay-windrowed	489	93.07	77.59	94.69	91.78	89.11	93.36
Oats	20	95.76	65.02	98.60	94.96	87.51	95.78
Soybeans-no till	968	58.45	51.00	60.27	56.88	64.74	56.82
Soybeans-min till	2468	55.77	45.97	47.35	54.45	53.86	45.20
Soybeans-clean till	614	72.67	46.25	67.50	72.61	52.83	67.55
Wheat	212	98.88	91.22	99.65	98.95	93.91	99.58
Woods	1294	87.52	69.90	95.19	87.07	80.55	92.06
Bldg-Grass-Tree-Drives	380	63.84	45.22	30.83	61.94	52.30	37.22
Stone-steel towers	95	88.75	93.13	92.11	89.68	94.50	91.82
Overall accuracy		68.47	53.22	66.51	67.84	63.29	65.09
Average accuracy		78.13	64.17	76.68	77.69	73.35	75.93
κ statistic		64.39	47.69	62.31	63.72	58.76	60.78

III. EXPERIMENTAL RESULTS

In this section, we evaluate the proposed SVMsub using a real hyperspectral image. A comparison with the conventional SVM, implemented with a Gaussian RBF kernel, is provided in order to evaluate the proposed formulation with regard to a widely used one in the remote sensing community. We emphasize that our selection of the RBF kernel has been done after extensive experiments with other types of kernels. In addition, we also perform comparisons with the MLRsub in [17]. Although it is well known that these methods can be combined with a spatial regularization technique such as the Markov random field [23] in order to enhance classification accuracies, we have decided to focus only on spectral-based classification in this contribution for clarity. We would also like to emphasize that, due to space considerations, we could not include a quantitative comparison to other widely used techniques for hyperspectral image classification. However, an extensive comparison of (classic and new) techniques for hyperspectral image classification has been recently presented in [24] and can be used as reference. The parameters of the different classification methods tested in this work have been carefully optimized by means of fivefold cross-validation. As for the parameters involved in the subspace estimation stage, we have also carefully optimized them following the procedure described in [17]. In all our experiments, we randomly select training samples from the reference data and report the overall average individual classification accuracies and the κ statistic obtained after 50 Monte Carlo runs.

The AVIRIS Indian Pines scene used in our experiments was collected over northwestern Indiana in June of 1992. The image is of size 145×145 pixels, with 220 spectral bands in the spectral range from 0.4 to 2.5 μm and nominal spectral resolution of 10 nm. In this letter, we use two versions of the hyperspectral data, one with all the 220 spectral bands available and another one with 200 channels after removal of 20 bands with noise and water absorption phenomena (this is intended to illustrate the performance of the proposed method in noisy scenarios). The reference information available for this scene comprises 16 mutually exclusive classes, with a total of 10 366 labeled samples. The scene has low spatial resolution, with a pixel size of 20 m. The scene represents a challenging classification scenario due to the spectral similarity between the classes.

In our first experiment with the AVIRIS Indian Pines scene, we analyze the performance of the SVMsub method under different noise conditions. Table I shows the classification results obtained by the different methods tested using only 320 training samples (20 samples per class). It should be noted that this is a very limited number of labeled samples, which we have set on purpose in order to address the fact that it is very common in practical scenarios that limited training sets are available. As shown by Table I, the proposed SVMsub obtained the best classification results in both scenarios (i.e., with all bands and with the noisy bands removed). Furthermore, MLRsub is also quite robust in noisy conditions, while the classic SVM appears to be more sensitive to noise. In addition, it is worth noting that the results obtained by subspace-based classifiers are comparable (or even slightly better) in the case in which all spectral bands (including noisy ones) are considered. This is because, even with noisy bands, the class-dependent subspace can be better estimated as the dimensionality increases. This is an important observation, which reveals that subspace-based techniques are important in order to fully exploit the information present in the original hyperspectral data.

For illustrative purposes, Fig. 1 shows some of the obtained classification maps provided by the different tested methods for the AVIRIS Indian Pines scene. These maps correspond to one of the 50 Monte Carlo runs conducted in each case. As shown by Fig. 1, the classification accuracies provided by the SVMsub are significantly higher than those provided by the classic

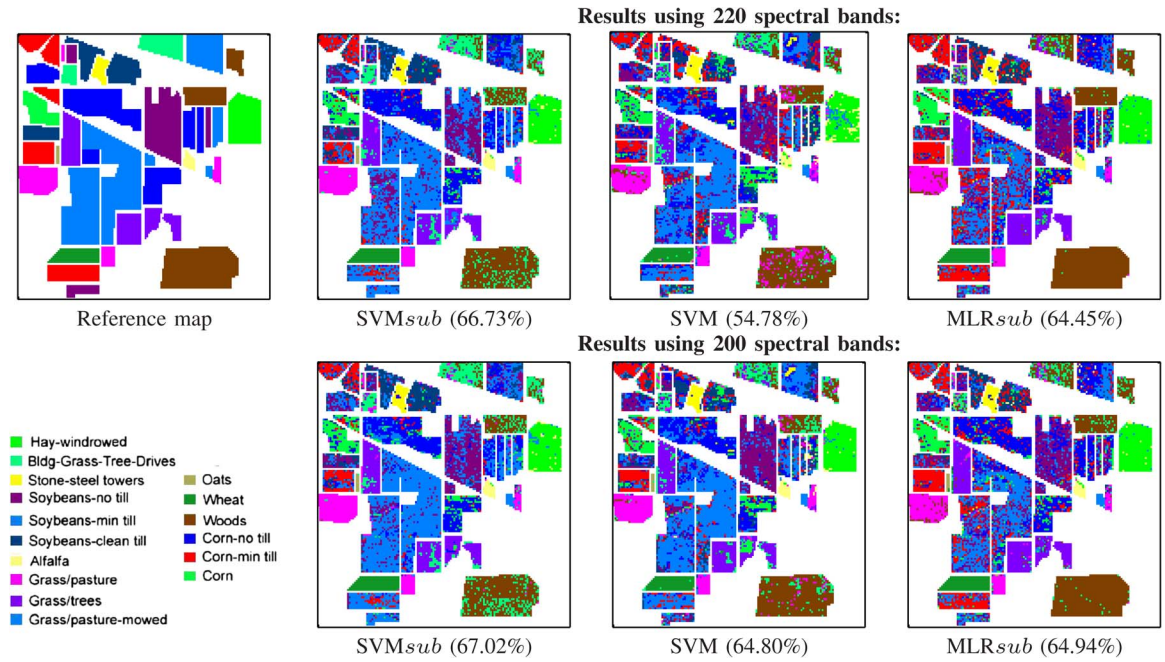

Fig. 1. Classification maps obtained by the different tested methods for the AVIRIS Indian Pines scene. In all cases, only 320 training samples in total (20 per class) were used. The overall classification accuracies are given in the parentheses.

TABLE II
OVERALL CLASSIFICATION ACCURACIES (IN PERCENT) (PLUS/MINUS THE STANDARD DEVIATION) AND κ STATISTIC (IN THE PARENTHESES) OBTAINED BY THE DIFFERENT TESTED METHODS FOR THE AVIRIS INDIAN PINES SCENE, USING DIFFERENT NUMBERS OF TRAINING SAMPLES. COMPUTATIONAL COST (INCLUDING BOTH TRAINING AND TESTING TIMES) IS ALSO INCLUDED. BOTH THE TOTAL NUMBER OF SAMPLES USED AND THE (APPROXIMATE) NUMBER OF TRAINING SAMPLES PER CLASS ARE GIVEN (IN THE PARENTHESES)

# Samples (per class)	Classification methods tested			
		SVMsub	SVM	MLRsub
160 (10)	Accuracy	61.16±2.82 (56.54)	39.23±3.91 (33.08)	61.37±2.52 (56.76)
	Time (seconds)	0.73	2.67	2.54
240 (15)	Accuracy	65.74±2.20 (61.45)	44.74±4.30 (39.15)	64.61±2.30 (60.24)
	Time (seconds)	0.77	3.44	3.27
320 (20)	Accuracy	68.47±2.06 (63.83)	53.22±2.15 (47.86)	66.51±1.85 (62.42)
	Time (seconds)	0.79	4.40	3.08
400 (25)	Accuracy	70.28±2.10 (66.36)	56.24±1.67 (50.89)	67.62±1.72 (63.46)
	Time (seconds)	0.81	5.51	3.30
480 (30)	Accuracy	72.01±1.57 (68.22)	59.03±1.73 (53.88)	69.29±1.81 (65.28)
	Time (seconds)	0.83	6.78	3.50
560 (35)	Accuracy	72.80±1.32 (69.08)	60.69±1.40 (55.64)	69.98±1.41 (66.02)
	Time (seconds)	0.85	8.07	3.82
640 (40)	Accuracy	73.69±1.66 (70.06)	62.05±1.00 (57.15)	71.04±1.35 (67.11)
	Time (seconds)	0.89	9.57	3.90
720 (45)	Accuracy	74.00±1.46 (70.37)	63.25±1.16 (58.44)	71.65±1.49 (67.73)
	Time (seconds)	0.91	11.53	4.22
800 (50)	Accuracy	74.36±1.58 (70.77)	64.16±0.96 (59.44)	71.47±1.34 (67.57)
	Time (seconds)	0.96	14.24	4.48

SVM formulation. Also, we emphasize that these results were obtained with a very limited number of training samples, which indicates that SVMsub can properly deal with the imbalance between the high dimensionality of the input data and the limited availability of training information. The proposed method can also deal with noise and mixed pixels, which dominate the AVIRIS Indian Pines since the agricultural features were very early in their growth cycle when the image was acquired.

In our second experiment with the AVIRIS Indian Pines scene, we evaluate the performance of the compared methods using different numbers of training samples. Here, we focus on the results obtained for the full hyperspectral image (with 220 spectral bands), as this is generally a more difficult problem due to noise and higher data dimensionality. Table II shows the overall classification accuracies (indicating the standard deviation) and the κ statistic obtained by the different methods

tested, as a function of the number of training samples. The computational cost, including both training and testing times, is also reported. In our experiments, we approximately used the same number of training samples per class (except for those classes which are very small). As shown by Table II, the proposed SVMsub obtained better results than the other tested methods in all cases. For instance, when a total of 560 labeled samples were used (approximately 35 samples per class), the proposed SVMsub obtained an overall accuracy of 78.92%, which is 19.23% higher than the one obtained by the traditional SVM and 8.94% higher than the one obtained by the MLRsub. Another important observation is that the subspace-based methods, i.e., SVMsub and MLRsub, are very fast even using a large number of training samples, while the conventional SVM method needs much more time when the number of training samples increases.

IV. Conclusion and Future Lines

In this letter, we have developed a new subspace-based SVM classifier called SVMsub. The main innovation of this classifier is that, in the construction of the SVM nonlinear function, we learn the subspace associated to each class. This formulation allows us to better cope with several phenomena that are quite important in hyperspectral image classification, such as the imbalance between the (high) dimensionality of the input data and the (limited) availability of training samples, as well as with the presence of noise and mixed pixels in the input data. The proposed method has been compared with the classic SVM formulation and also with a previously developed subspace-based technique based on the multinomial logistic regression classifier (MLRsub), obtaining good classification results with very limited training samples. In future work, we will explore the impact of including spatial information in the proposed formulation and also test the method under different conditions and analysis case studies.

Acknowledgment

The authors would like to thank the Associate Editor and the two anonymous Reviewers for their detailed and highly constructive criticisms, which have greatly helped them improve the quality and presentation of their manuscript.

References

[1] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral–spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, Mar. 2013.

[2] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*. New York, NY, USA: Wiley, 2003.

[3] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis: An Introduction*. Berlin, Germany: Springer-Verlag, 2006.

[4] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, no. S1, pp. 110–122, Sep. 2009.

[5] J. Chanussot, M. M. Crawford, and B.-C. Kuo, "Foreword to the special issue on hyperspectral image and signal processing," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 3871–3876, Nov. 2010.

[6] J. Bioucas-Dias and J. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.

[7] A. Plaza, Q. Du, J. Bioucas-Dias, X. Jia, and F. Kruse, "Foreword to the special issue on spectral unmixing of remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4103–4110, Nov. 2011.

[8] J. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[9] A. Plaza, P. Martinez, R. Perez, and J. Plaza, "A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 650–663, Mar. 2004.

[10] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.

[11] J. J. Settle, "On the relationship between spectral unmixing and subspace projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 4, pp. 1045–1046, Jul. 1996.

[12] J. C. Harsanyi and C.-I. Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 4, pp. 779–785, Jul. 1994.

[13] C.-I. Chang, "Orthogonal Subspace Projection (OSP) revisited: A comprehensive study and analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 502–518, Mar. 2005.

[14] C. Chang, X. Zhao, M. L. G. Althouse, and J. J. Pan, "Least squares subspace projection approach to mixed pixel classification for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 3, pp. 898–912, May 1998.

[15] R. Larsen, M. Arngren, P. Hansen, and A. Nielsen, "Kernel based subspace projection of near infrared hyperspectral images of maize kernels," in *Image Analysis*. Berlin, Germany: Springer-Verlag, 2009, pp. 560–569.

[16] M. D. N. Acito and G. Corsini, "Hyperspectral signal subspace identification in the presence of rare signal components," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 4, pp. 1940–1954, Apr. 2010.

[17] 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, Mar. 2012.

[18] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[19] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.

[20] B. Scholkopf and A. Smola, *Learning With Kernels—Support Vector Machines, Regularization, Optimization and Beyond*. Cambridge, MA, USA: MIT Press, 2002, ser. MIT Press Series.

[21] V. Vapnik, *Statistical Learning Theory*. New York, NY, USA: Wiley, 1998.

[22] B. Scholkopf, A. J. Smola, and K.-R. Muller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, Jul. 1998.

[23] M. C. Zhang, R. M. Haralick, and J. B. Campbell, "Multispectral image context classification using stochastic relaxation," *IEEE Trans. Syst., Man, Cybern.*, vol. 20, no. 1, pp. 128–140, Jan./Feb. 1990.

[24] J. Bioucas-Dias *et al.*, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.