Spectral-Spatial Classification of Hyperspectral Images Using ICA and Edge-Preserving Filter via an Ensemble Strategy

Junshi Xia, Lionel Bombrun, Tülay Adali, Yannick Berthoumieu, Christian Germain

HAL Id: hal-01379723
https://hal.science/hal-01379723

Spectral-Spatial Classification of Hyperspectral Images using ICA and Edge Preserving Filter via an Ensemble Strategy

Junshi Xia, *Member, IEEE,* Lionel Bombrun, Tülay Adalı, *Fellow, IEEE,* Yannick Berthoumieu, *Member, IEEE,* and Christian Germain

Abstract

In order to obtain accurate classification results of hyperspectral images, both the spectral and spatial information should be fully exploited in the classification process. In this paper, we propose a novel method using independent component analysis (ICA) and edge-preserving filtering (EPF) via an ensemble strategy for the classification of hyperspectral data. First, several subsets are randomly selected from the original feature space. Second, ICA is used to extract spectrally independent components followed by an effective EPF method, to produce spatial features. Two strategies (i.e., parallel and concatenated) are presented to include the spatial features in the analysis. The spectral-spatial features are then classified with a random forest (RF) or rotation forest (RoF) classifier. Experimental results on two real hyperspectral datasets demonstrate the effectiveness of the proposed methods. A sensitivity analysis of the new classifiers is also performed.

Index Terms

Classification, hyperspectral data, independent component analysis (ICA), edge preserving filter (EPF).

I. INTRODUCTION

During the past two decades, the development of hyperspectral sensors have resulted in great improvement for the image acquisition capabilities. Hyperspectral sensors are now able to provide images with both high spectral and spatial resolutions. Hence, hyperspectral data offers a unique opportunity to monitor the Earth surface. Thematic applications include environmental mapping and crop monitoring [1], [2].

Supervised classification is one of the most important problems in the remote sensing community. Given a set of training samples (i.e., pixel vectors for hyperspectral image), the aim of classification is to assign a unique

class label to each pixel after a training process. Classical techniques, such as maximum likelihood and nearest neighbor classifiers, can be used to perform a supervised classification of hyperspectral data. Nevertheless, their classification accuracy is relatively low because a sufficient number of training samples is often not available for the wide range of spectral bands. This often causes an unbalance between the high dimensionality of spectral bands and the limited number of training samples, known as the curse of dimensionality (also referred to as the Hughes phenomenon) [3]. In addition, the spectral bands of hyperspectral data are highly correlated and might show some redundancy. Furthermore, due to the spatial variability observed for high-resolution data, classification algorithms exploiting only the spectral information generally demonstrate a low performance. To overcome this, spatial contextual information (relationship between neighboring pixels) should be included in the analysis, resulting in *spectral-spatial* classifiers [4], [5]. In recent years, many spectral and/or spatial classification methods have been developed to alleviate the aforementioned issues, i.e., high dimensionality of the feature space, correlated features and spatial variability, in order to improve the classification performance of hyperspectral data.

A popular strategy for providing enhanced classification performance is the random subspace (RS) ensemble [6]. The idea is intuitive and simple: subsets of feature sets are used in the ensemble instead of using all features. Each classifier in the ensemble is constructed on a different feature subset by randomly sampling the original feature set. The rationale behind the RS ensemble is to break down a complex high-dimensional problem into several lower-dimensional sub-problems, hence allowing to address the curse of dimensionality problem [7].

Discriminative classifiers, e.g., support vector machines (SVMs) [8]–[10], random forest (RF) [11], [12] and rotation forest (RoF) classifiers [13], [14], are the most widely used pixel-wise classifier for hyperspectral data. Those classifiers provide good performances in terms of classification accuracies. In [14], Xia *et al.* propose RoF for the classification of hyperspectral data and note its superior performance to RF and completion with the SVMs. Compared to the SVMs, the two main advantages of RF and RoF, i.e., low computational complexity and few parameters to tune, motivate us to select them as the base classifiers in this work. However, to improve the classification performance of RF and RoF, the input features should be independent [15], [16].

In order to obtain independent features, a feature extraction (FE) step is generally employed. Principal component analysis (PCA) is one of the most frequently used FE method in the remote sensing community [17]. Recently, independent component analysis (ICA) has received attention for FE of hyperspectral remote sensing images [18]–[21]. In particular, ICA extracts underlying source components that give rise to the mixed signal measured by the sensor and the informative components present in the scene [18], [21]. Classically, when ICA is employed to hyperspectral data, PCA is performed first, and then, ICA is performed on a percentage of the top most important principal components, while the remaining components are discarded [21]–[23]. However, the PCA process might lead to a loss of useful information in the discarded components, which may contain discriminant information to improve the classification accuracy. In [24], it has been suggested that a noise-adjusted principal component analysis (NAPCA) offers more useful information of objects in the original data than those obtained from PCA. In order to enhance the classification performance by using ICA without losing useful information in the PCA process, we propose to use a random subspace ensemble approach, in which several subsets are randomly selected from

the original spectral bands and then ICA is applied on each subset. Furthermore, the classification result can be improved by the integration of spatial contextual information.

Previous spectral-spatial classifiers include spatial contextual information, such as mathematical morphology (MM) and Markov random fields (MRFs). Li *et al.* [25] have developed a generalized composite kernel (GCK) framework by combining spectral information and the most effective MM features: extended multi-attribute profiles (EMAPs). Falco et al. [26] proposed a spectral-spatial classifier based on ICA and reduced attribute profiles. Recently, Xia *et al.* have proposed two powerful classifiers, namely rotation forests with MRFs (RoF-MRF) [27] and random subspace ensembles with EMAPs (RS-EMAPs) [16]. The former combines the class posterior probabilities produced by RoF and the spatial information represented by MRF-based multilevel logistic (MLL) prior [27]. The latter uses random subspace ensembles to classify EMAPs features [16]. Recently, edge-preserving filters (EPF) have been successfully applied in many fields such as denoising [28]. Those filters are used to remove noise, weak edges, and small details whereas the overall structure of the image is preserved. Kang *et al.* have successfully combined a probabilistic SVM with an EPF for the classification of hyperspectral images [29]. Furthermore, they proposed to extract spatial features of hyperspectral images with image fusion and recursive filtering techniques (IFRF) [30].

In order to tackle the three main issues of the classification of hyperspectral images with both high spectral and high spatial resolutions (i.e., high dimensionality of the feature space, correlated features and spatial variability), we present a new spectral-spatial classification scheme. First, several subsets are randomly selected from the original feature space to reduce the dimensionality. Second, ICA is used to extract independent features. Third, an EPF is used to reduce the spatial variability in the feature set. Parallel ensemble and concatenated models are finally used to include those spatial features into an RF or RoF classifier. The novelty of this work consists of:

- a new spectral-spatial classification scheme by using ICA and EPF via an ensemble strategy;
- introducing two novel ICA techniques, i.e., entropy bound minimization (EBM) [31] and entropy rate bound minimization (ERBM) [32], [33], to the hyperspectral remote sensing community;
- exploiting the spatial contextual information by means of a recent and effective EPF, rolling guidance filter (RGF).

In this study, previous work that introduces such a parallel ensemble [34] is extended, proposing a concatenated model to supervised classification based on ICA and EPF. In addition, the new methodology is also validated on additional hyperspectral dataset.

The paper is organized as follows. Section II recalls the principle of ICA while Section III gives a brief overview of EPF. Section IV is devoted to the introduction of the proposed spectral-spatial classifier. Sections V and VI present some experimental results on two hyperspectral dataset. Conclusions and future work are finally reported in Section VII.

II. INDEPENDENT COMPONENT ANALYSIS (ICA)

ICA is an attractive solution to the blind source separation (BSS) problem, which decomposes an observed set of mixtures into a set of statistically independent components (ICs) [18]. We consider the observed mixture

$\mathbf{x} = [x_1, ..., x_D]^T = \mathbf{As}$, which can be viewed as a linear combination of D random variables—or random processes when sample dependence is taken into account—$\mathbf{s} = [s_1, ..., s_n]^T$ through a $D \times D$ non-singular mixing matrix $\mathbf{A}$. The statistically ICs $\mathbf{y} = [y_1, ..., y_D]^T$ are estimated by forming $\mathbf{y} = \mathbf{Wx}$, where $\mathbf{W}$ is the estimated demixing matrix that makes use of some form of diversity. The most widely used type of diversity is non-Gaussianity, i.e., higher-order-statistics (HOS). Two popular ICA approaches used in the remote sensing community are FastICA [35] and Joint Approximate Diagonalization of Eigenmatrices (JADE) [36]. FastICA [35] uses a fixed nonlinearity to maximize the non-Gaussianity while JADE [36] extracts the demixing matrix $\mathbf{W}$ by joint a diagonalization of the fourth-order cumulant matrix.

A more attractive approach to make use of HOS is to use a dynamic nonlinearity that is matched to each of the estimated source densities, y_d, for $d = 1, \ldots, D$ separately. Entropy bound minimization (EBM) utilizes an efficient entropy estimator to approximate the density of the sources by maximizing the entropy bound and using a finite number of prespecified measuring functions [31]. It provides robust performance according to the four measuring functions proposed in [31] but also allows for selection of nonlinearities using prior information about the sources. Another important type of diversity, which is of particular interest here, is sample dependence. The adjacent pixels in an image are highly correlated and making use of this additional statistical property promises to further improve the performance of ICA [37]. Entropy rate bound minimization (ERBM) [32], [33] effectively combines the dynamic nonlinearity selection of EBM with an invertible filter model and hence achieves better performance in terms of minimization of mutual information rate

$$\mathcal{I}_r(y_1; ...; y_D) = \sum_{i=1}^{D} H_r(y_i) - \log|\det(\mathbf{W})| - H_r(\mathbf{x}) \tag{1}$$

where, $H_r(y_i)$ is the (differential) entropy rate of the process y_i and $H_r(\mathbf{x})$ is a constant with respect to $\mathbf{W}$. For EBM, we can consider the same cost function where the entropy rate is simply replaced by the entropy. When the demixing matrix $\mathbf{W}$ is assumed to be orthogonal and the nonlinearity that corresponds to source distribution is fixed, we obtain FastICA [38].

III. Edge preserving filter (EPF)

EPF is an image processing technique that aims to reduce the spatial variability. It smooths away textures whilst retaining sharp edges [28]. For high spatial resolution hyperspectral images, the neighboring pixels usually have strong relationships with each other. The use of EPF makes that the neighboring pixels on the same side of an edge have similar features values, which is beneficial for improving the classification performance. In this paper, we propose to use one of the most recent and effective EPF filters, the RGF.

RGF, which is based on a modification of the bilateral filter [39], effectively removes noise and small details while preserving large-scale structures automatically, which the standard bilateral filter often fails to do [40]. It is composed of two steps, i.e., small structure removal and edge recovery. Small structure removal can be done by a Gaussian filtering. Then, a joint bilateral filter is used to recover the edge iteratively. The result of the $t-$th iteration is denoted by J^{t+1}. J^1 is initially set as the output of the Gaussian filtering. J^{t+1} is calculated by a joint bilateral

filtering form given the input image I and the previous iteration result J^t.

$$J^{t+1}(i) = \frac{1}{Q_i} \sum_{j \in \mathcal{N}_p} \exp\left(-\frac{\|i-j\|^2}{2\sigma_s^2} - \frac{\|J^t(i) - J^t(j)\|^2}{2\sigma_r^2}\right) I(j) \tag{2}$$

where, $Q_i = \sum_{j \in \mathcal{N}_p} \exp\left(-\frac{\|i-j\|^2}{2\sigma_s^2} - \frac{\|J^t(i) - J^t(j)\|^2}{2\sigma_r^2}\right)$. The indexes i and j are the coordinates of pixels in the image, $\mathcal{N}_p$ denotes the set of neighbor pixels of i, σ_s and σ_r control the spatial and range weights respectively.

In general, those steps can be combined into one by starting the rolling guidance simply from a constant-value (equal to C) image. In this case, the initial step can be saved by starting rolling guidance from J^0, where $\forall i$, $J^0(i) = C$. Algorithm 1 depicts the RGF implementation.

Algorithm 1 Rolling Guidance Filter

Input: I, σ_s, σ_r and n^{iter}
 1: Initialize J^0 as a constant image
 2: For $i = 1 : n^{\text{iter}}$
 3: $J^t \leftarrow \text{JointBilateral}(I, J^{t-1}, \sigma_s, \sigma_r) \triangleright$ using (2)
 4: Endfor
Output: The output image $G \leftarrow J^{n^{\text{iter}}}$

IV. Classification using ICA and EPF via an ensemble strategy

In this paper, we present two strategies to combine multiple features obtained for ICA and RGF. The first one is the parallel combination (referred as E-ICA-RGF$_P$) that is based on the separate classification of each feature group and on the fusion of the results obtained by the independent classifiers in order to generate the final decision result. The second one is the concatenated combination (referred as E-ICA-RGF$_C$), in which multiple features are integrated into one vector and then classified via a classifier. In comparison to the concatenation combination, the parallel combination keeps the dimensionality of the data low and increases the robustness of the results, particularly if different features generate classification results with sufficient diversity. However, even if complementary information can be extracted by considering different features, a great redundancy is present in the extracted features. Thus, it is advisable that a classification algorithm with excellent penalization capability is used for classifying the features in order to handle the increased dimensionality which can lead to the curse of dimensionality.

As shown in Fig. 1, the proposed methods (E-ICA-RGF$_P$ and E-ICA-RGF$_C$) consist of five steps:

- *Step 1.* Random subspace ensemble (E): number of subsets is set to K and M features in each subset are randomly selected without replacement from original spectral bands.

- *Step 2.* ICA: the aim of this step is to extract informative ICs for the classes in each subset, to be used for classification. Here, we propose to use EBM and ERBM due to their superior separation performance [31]–[33].

- *Step 3.* RGF: RGF is performed on each extracted component to obtain the mth feature in ith subset.

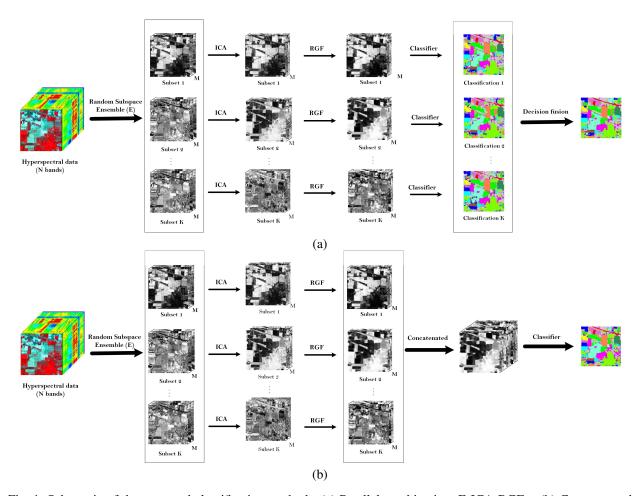$$G_k^m \leftarrow \text{Algorithm 1}\left(IC_k^m, \sigma_s, \sigma_r, n^{\text{iter}}\right) \tag{3}$$

Fig. 1: Schematic of the proposed classification methods. (a) Parallel combination: E-ICA-RGF$_P$. (b) Concatenated combination: E-ICA-RGF$_C$.

where σ_s and σ_r are the spatial and range standard deviations of the filter in (2). IC_k^m is the mth ($m = 1, ..., M$) component derived from ICA in the kth ($k = 1, ..., K$) subset. G_k^m is the resulting feature obtained after RGF.

Parallel combination (P)

- *Step 4*: perform classification on the filtered images in each subset. Two classifiers (e.g., RF and RoF) are used. For RoF, the feature space is randomly split into several subsets. Then, principal component analysis (PCA) is applied on each subset. Furthermore, a new training data is formed by concatenating all the principal components in each subset. An individual decision tree (DT) classifier is trained with this training set. A series of individual classifiers is generated by repeating the above steps several times. The final classification result is produced by integrating the results from individual classifiers using a majority voting rule [13], [27], [41]. For both RF and RoF, two parameters, i.e., the number of trees and the number of selected features, are empirically fixed. According to the study presented in [16], the number of trees of the RF and RoF classifiers are set to 100 and 20 respectively. Moreover, the number of selected features is set to $\sqrt{M}$.

- *Step 5*: combine the results together to generate the final classification map by a majority vote rule. The rule

TABLE I: Indian Pines AVIRIS and University of Pavia ROSIS image: class name and number of samples in reference

	AVIRIS		ROSIS	
No.	Name	Reference data	Name	Reference data
1	Alfalfa	54	Bricks	3682
2	Corn-no till	1434	Shadows	947
3	Corn-min till	834	Metal Sheets	1345
4	Bldg-Grass-Tree-Drives	234	Bare Soil	5029
5	Grass/pasture	497	Trees	3064
6	Grass/trees	747	Meadows	18649
7	Grass/pasture-mowed	26	Gravel	2099
8	Corn	489	Asphalt	6631
9	Oats	20	Bitumen	1330
10	Soybeans-no till	968		
11	Soybeans-min till	2468		
12	Soybeans-clean till	614		
13	Wheat	212		
14	Woods	1294		
15	Hay-windrowed	380		
16	Stone-steel towers	95		

considered when combining the results of the single classifiers relies on the sum of the votes of the classifiers applied to the features obtained from ICA and RGF, assigning each pixel to a class.

Concatenated combination (C)

- *Step 4*: concatenate the features together.

- *Step 5*: perform classification on the concatenated features to produce the final result (the same with step 4 of parallel combination).

V. DATASETS AND SETUP FOR EXPERIMENTAL RESULTS

A. Hyperspectral datasets

Two different real hyperspectral data are used to evaluate the performances of the proposed approaches. The two hyperspectral images provide different characteristics in terms of spatial and spectral resolutions in order to validate the methods in very different scenarios. The scenes have the following characteristics:

1) Indian Pines AVIRIS: The first hyperspectral image is recorded by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over the Indian Pines in Northwestern Indiana, USA. This scene, which comprises 220 spectral bands in the wavelength range from 0.4 to 2.5 μm with a spectral resolution of 10 nm, is composed of 145×145 pixels, and the spatial resolution is 20 m/pixel. The reference data with sixteen classes of interest is composed of 10336 pixels (seen in Table I). Figs. 2(a) and (b) show respectively the three-band color composite image and the ground truth of AVIRIS hyperspectral data.

2) University of Pavia ROSIS: The second experiment was carried out on the University of Pavia image of an urban area operated by Reflective Optics Spectrographic Imaging System (ROSIS)-03 optical airborne sensor. The original image is composed by 610×340 pixels, with a very high spatial resolution of 1.3 m/pixel and 115 spectral bands. In this work, 12 noisy channels were removed and the remaining 103 spectral bands are used for the investigation [42]. Nine land cover classes were considered for classification (seen in Table I). Fig. 3 shows the three-band color composite image and reference map of University of Pavia data.
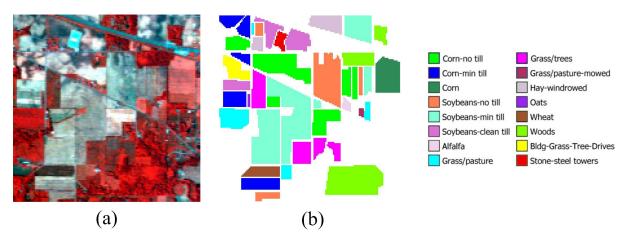
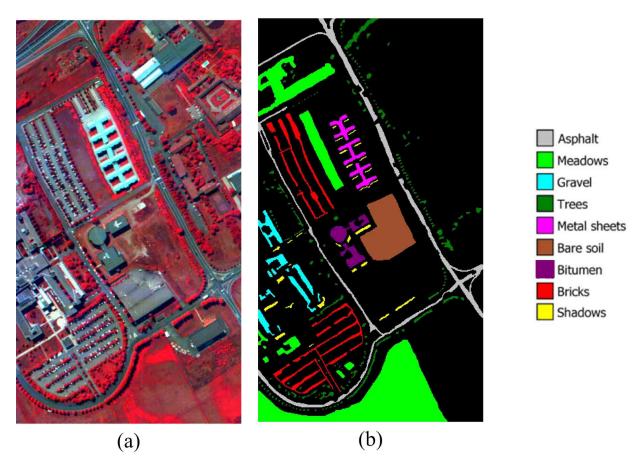Fig. 2: (a) Three-band color composite of AVIRIS image. (b) Ground truth.

Fig. 3: (a) Three-band color composite image of AVIRIS data. (b) Reference map.

B. Experimental setup

We conducted several experiments with these hyperspectral images in order to investigate several relevant aspects of our proposed methods, such as spectral analysis, spectral-spatial analysis and parameter sensitivity analysis. A small set of labeled samples (30 samples per class) are randomly selected from the reference data as the training set. The rest of the pixels forms the testing set [1]. In order to increase the statistical significance of the results, we perform ten Monte Carlo runs with a set of different training samples each run. Experiments for the two hyperspectral datasets consist in:

- Spectral analysis: we present the classification results for the proposed parallel method without the filtering step (E-ICA) and compare it with the ICA applied to the lower-dimensional space (PCA as the dimensionality reduction technique) and high-dimensional space (entire dataset). In this case, K is set to 10, and M is set equal to the number of classes. For this case, when ICA is applied to the lower-dimensional space (referred as PCA-ICA), we first use PCA on the original space and then apply ICA to the obtained feature subset. The procedure is repeated for each ICA algorithm, resulting in different subsets of the kept components, starting from a minimum of 5 components up to 40 components. Only the best results are reported. For the strategy of high-dimensional space, ICA is applied to the entire data set and then the most informative features are selected by using the Relief method [43]. We do not consider JADE and ERBM in this case, since their computational load increase significantly with dimensionality.

- Spectral-spatial analysis: we present results obtained from the parallel and concatenated combinations: E-ICA-RGF$_P$ and E-ICA-RGF$_C$, and compare them with the following algorithms: 1) random subspace ensemble (E): proposed method without ICA and RGF; 2) random subspace ensemble with ICA (E-ICA): proposed method without RGF and 3) random subspace ensemble with RGF (E-RGF): proposed method without ICA. To evaluate the influence of the final classifier, a comparison between and RF and RoF is shown. It should be noted that E, E-ICA and E-RGF are constructed in a parallel way, and σ_s and σ_r in RGF are set to 7 and 0.1, respectively.

- Parameter sensitivity analysis. We evaluate the influences of K, M, σ_s and σ_r on classification performances.

- Comparisons with other state-of-the art classifiers. Five state-of-the-art spectral-spatial classifiers, such as generalized composite kernels (GCK) [25], rotation forest with Markov random fields (RoF-MRF) [27], random subspace with extended morphological attribute profiles (RS-EMAPs) [16], the SVMs with edge preserving filtering (SVM-EPF) [29] and image fusion and recursive filtering (IFRF) [30], are added for comparison against the proposed methods. The settings of these methods can be found in the original references.

The following measures are used to evaluate the performances:

- Overall accuracy (OA): the percentage of correctly classified samples; and

- Average accuracy (AA): average percentage of correctly classified samples for individual class.

---

[1]For the minority class *Grass/pasture-mowed* and *Oats*, we select half of the samples for training (10 and 13 samples respectively).

TABLE II: Indian Pines *AVIRIS* image. Overall and average accuracies obtained from the proposed E-ICA method in comparison to the ICA applied on the lower-dimensional space (PCA as dimensionality reduction) and high-dimensional space (entire dataset).

Methods		RF		RoF	
		OA	AA	OA	AA
Original		61.60±1.33	71.39±0.72	66.46±1.28	75.59±0.75
Entire data-ICA	EBM	**58.44±2.54**	**70.12±1.85**	**61.08±2.17**	**71.14±1.41**
	FastICA	35.14±3.46	54.12±3.12	41.64±3.78	62.37±2.98
PCA-ICA	EBM	65.91±1.71	76.46±1.61	67.58±1.65	77.92±1.43
	ERBM	**68.69±1.65**	**78.23±1.46**	**69.08±1.67**	**78.86±1.38**
	JADE	60.95±1.81	72.25±1.24	61.18±1.74	72.91±1.61
	FastICA	64.36±1.69	75.41±1.51	62.04±1.81	72.57±1.64
E-ICA	EBM	68.82±1.62	79.33±1.50	68.02±1.21	78.28±0.81
	ERBM	**72.61±1.27**	**83.55±1.15**	**69.27±1.38**	**81.11±0.96**
	JADE	65.36±1.82	71.67±1.62	61.54±1.64	70.49±1.21
	FastICA	65.29±1.65	72.27±1.57	61.54±1.79	69.19±1.46

VI. EXPERIMENTAL RESULTS AND DISCUSSION

A. Spectral analysis

We test the proposed E-ICA approach for spectral analysis and show the results on the two datasets. Tables II and III show the OAs and AAs obtained from the E-ICA method in comparison to ICA applied to the lower- and high-dimensional spaces for the AVIRIS and ROSIS images. From Tables II and III, it can be found that the average OA (AA) obtained for RF and RoF classifiers when applied to the original feature space of AVIRIS image are 61.60% (71.39%) and 66.46% (75.59%), respectively. The average OAs (AAs) achieved by RF and RoF classifiers when applied to the original feature space of ROSIS image are 69.18% (78.53%) and 75.62% (83.41%), respectively. As observed in these two tables, the ICA results for the high-dimensional space do not improve the performances as the selected components in this case tend to be noisy, resulting in low accuracy classification. On the contrary, PCA-ICA approach leads to significant increase in the classification accuracy. This indicates that the pre-processing step (e.g., dimensionality reduction) helps ICA to provide better features for obtaining accurate classification results. These results are consistent with the study presented in [21]. The proposed E-ICA method significantly increases the classification accuracy in comparison to the other two strategies since it adopts an ensemble strategy to combine the results obtained for ICA on the lower dimensional spaces (the subsets) while retaining spectral information. In particular, the two ICA algorithms, EBM and ERBM, yield better results and lower standard deviations than JADE and FastICA. Indeed, ERBM not only matches a wide range of distributions like EBM [31] but also considers sample dependence [32], [33].

B. Spectral-spatial analysis

In this section, the results obtained for the proposed spectral-spatial classifiers, E-ICA-RGF$_P$ and E-ICA-RGF$_C$ are presented. With the help of RGF, E-ICA-RGF$_P$ and E-ICA-RGF$_C$ significantly improve the classification results compared with the classifiers which consider only spectral information (seen in Section VI.A). We would like to

TABLE III: University of Pavia *ROSIS* image. Overall and average accuracies obtained from the proposed E-ICA method in comparison to the ICA applied on the lower-dimensional space (PCA as dimensionality reduction) and high-dimensional space (entire dataset).

Methods		RF		RoF	
		OA	AA	OA	AA
Original		69.18±2.74	78.53±0.93	75.62±2.14	83.41±0.89
Entire data-ICA	EBM	**65.17±2.61**	**74.12±1.01**	**67.25±1.81**	**75.14±0.92**
	FastICA	58.53±3.12	68.42±1.42	61.64±3.45	74.35±1.58
PCA-ICA	EBM	75.68±2.11	83.19±1.01	73.92±1.95	82.70±0.86
	ERBM	**76.98±2.12**	**83.59±1.06**	**77.34±1.86**	**85.86±0.81**
	JADE	71.93±2.45	78.96±1.24	69.89±2.01	79.50±1.01
	FastICA	74.06±2.32	80.87±1.14	72.47±2.06	79.17±1.13
E-ICA	EBM	76.61±2.00	84.46±0.77	76.37±1.73	84.05±0.68
	ERBM	**77.12±1.91**	**85.31±0.78**	**77.89±1.81**	**85.10±0.71**
	JADE	75.42±2.08	82.43±0.81	75.61±1.96	82.25±0.78
	FastICA	73.97±2.16	80.43±0.91	77.14±1.91	81.58±0.83

emphasize that in this work we performed E-ICA-RGF$_P$ and E-ICA-RGF$_C$ with different ICA methods (e.g., EBM, ERBM, FastICA and JADE) and found that E-ICA-RGF$_P$ and E-ICA-RGF$_C$ with EBM and ERM perform slightly better than the ones with other ICA methods. The main weakness of ERBM is high computational complexity. Thus, considering the balance of classification performance and computational complexity, EBM provides a good trade-off. In the following, we only present results obtained from the proposed spectral and spatial methods using EBM.

Tables IV and V report the classification accuracies achieved for the proposed methods as well as other compared methods using RF and RoF classifiers, respectively (Indian Pines AVIRIS image). Classification accuracies of University of Pavia ROSIS image are shown in Tables VI and VII. Figs. 4 and 5 provide the classification maps (one sample out of the ten Monte Carlo runs). It can be seen that random subspace ensemble method does not improve the performance. The main reason is that we select a small number of features in each subset. For *AVIRIS* image, E-ICA and E-RGF produce higher OAs and AAs than RS and original, indicating the effectiveness of ICA and RGF techniques. E-RGF is superior to E-ICA for this dataset. For ROSIS image, the classification accuracy for classes 2, 5 and 8 decrease, leading to lower AA of E-RGF than the one of original E-ICA. This is due to the fact that when RGF is directly applied to the original feature space, it neglects the informative parts with small structures, such as shadow (class 2), trees (class 5) and asphalt (class 8). The ensemble strategy that combines the results obtained for ICA can extract the informative class-specific features, even for classes with small-scale objects. Followed by RGF, it ensures that the neighboring pixels belonging to the same class have similar feature values, thus decreasing the variability within regions belonging to the same class. This explains why the proposed methods have shown significantly better performances than E-RGF (see McNemar's test [2] in Table VIII).

In this case, the classification results obtained for the parallel combination (AVIRIS image) achieved 31.6, 31.5,

---

[2]McNemar's test (Z) is calculated by $Z = \frac{f_{12}-f_{21}}{\sqrt{f_{12}+f_{21}}}$, where, f_{12} means the number of samples correctly classified by classifier 1 and incorrectly classified by classifier 2. The difference between classifiers 1 and 2 is to be statistically significant if $|Z| > 1.96$. $Z > 0$ indicates that classifier 1 is more accurate than classifier 2.

TABLE IV: Indian Pines *AVIRIS* image. Classification results achieved by the RF classifier. For each method, "OA (%)", "AA (%)" and class-specific accuracies, "CA (%)" are reported.

Class	Original	E	E-ICA	E-RGF	E-ICA-RGF$_P$	E-ICA-RGF$_C$
1	81.25	85.00	90.00	96.67	98.75	**99.17**
2	40.16	41.84	66.80	68.75	90.70	**91.97**
3	48.54	50.19	58.69	64.83	89.47	**93.38**
4	67.60	70.69	81.13	95.25	**99.75**	99.26
5	83.49	82.68	82.46	78.61	93.53	**95.27**
6	79.61	80.85	95.54	83.53	**98.84**	98.08
7	87.69	89.23	86.92	94.62	**99.23**	98.46
8	86.54	84.47	96.56	96.12	**99.80**	99.63
9	77.00	76.00	81.00	83.00	**100.00**	**100.00**
10	64.37	65.82	72.75	66.52	91.86	**92.53**
11	52.25	50.96	40.14	59.62	**88.24**	87.60
12	48.99	48.46	72.24	69.14	**94.67**	94.43
13	94.62	94.34	**99.51**	94.45	99.34	99.23
14	83.31	83.41	92.99	82.48	**98.27**	97.60
15	48.83	41.31	56.54	93.57	**98.23**	97.54
16	98.00	98.00	96.00	98.92	**99.08**	98.92
OA	61.60±1.33	61.53±1.43	68.82±1.62	72.75±2.76	93.15±1.19	**93.43±0.99**
AA	71.39±0.72	71.45±0.83	79.33±1.50	82.88±1.72	**96.23±0.59**	96.23±0.60

24.3 and 20.4 percentage points over original, E, E-ICA and E-RGF, respectively. From Figs. 4 and 5, without RGF smoothing, the classification maps of the original, E and E-ICA look noisy due to the existence of mixed pixels. The classification maps obtained for methods involving spatial information (see Figs. 4 and 5(e)-(f)) show more homogeneous regions.

C. Parameter sensitivity analysis

In this part, we investigate the sensitivity of the proposed methods to parameter choice. Figs. 6 and 7 plot the OA as a function of the key parameters for the proposed E-ICA-RGF$_P$ (AVIRIS image) and E-ICA-RGF$_C$ (ROSIS image) with RF classifier. From Figs. 6 and 7, it is observed that: 1) There is no pattern of dependence between K and the accuracy. The OA of concatenated combination does not decrease as K increases, indicating that RF is not sensitive to the increased dimensionality; 2) When M increases, the proposed method tends to give better performance at the expense of increased computational complexity; 3) For AVIRIS image, the proposed method achieves the best classification performance when $\sigma_s = 7$ but the performance is satisfactory over a wide range of values as well. For the ROSIS image, the OA of the proposed method increases with σ_s; 4). The best range of σ_r for the two datasets is between 0.1 to 0.3. Hence, based on the results obtained from AVIRIS and ROSIS images, selection of parameters is not very critical for the proposed methodology, which is an important added advantage. The only parameter that seems to provide better performance in a small range is σ_r. In practice, the users might select a small value of σ_r as in our case to better preserve the edges of hyperspectral data and hence increase the discrimination between different classes.

TABLE V: Indian Pines *AVIRIS* image. Classification results achieved by the RoF classifier. For each method, "OA (%)", "AA (%)" and class-specific accuracies, "CA (%)" are reported.

Class	Original	E	E-ICA	E-RGF	E-ICA-RGF$_P$	E-ICA-RGF$_C$
1	82.50	84.58	90.42	96.25	**96.88**	96.67
2	54.49	50.75	76.57	75.73	**93.01**	90.33
3	57.67	61.63	57.10	76.53	**91.17**	88.72
4	80.83	79.12	81.47	94.71	**99.71**	98.38
5	85.05	84.43	87.69	85.93	**95.46**	93.75
6	84.20	86.92	96.22	86.56	**99.07**	98.12
7	80.00	85.38	87.69	98.46	99.23	**100.00**
8	84.14	90.11	93.33	97.06	**99.41**	98.95
9	83.00	87.00	63.00	95.00	**99.00**	**99.00**
10	69.89	72.63	69.50	64.22	**90.28**	89.63
11	45.44	52.91	32.68	58.40	82.78	**83.21**
12	63.32	55.17	71.59	73.07	92.64	**92.74**
13	96.43	95.44	**99.62**	92.86	99.34	97.31
14	86.26	86.55	92.16	79.83	**97.67**	95.83
15	61.57	51.89	59.51	86.97	97.54	**98.43**
16	94.62	**97.85**	94.00	97.54	96.77	96.77
OA	66.46±1.28	66.99±1.45	68.02±1.21	74.40±1.67	**91.97±0.92**	91.01±0.84
AA	75.59±0.75	76.40±0.89	78.28±0.81	84.94±0.94	**95.56±0.63**	94.86±0.61

TABLE VI: University of Pavia *ROSIS* image. Classification results achieved by the RF classifier. For each method, "OA (%)", "AA (%)" and class-specific accuracies, "CA (%)" are reported.

Class	Original	E	E-ICA	E-RGF	E-ICA-RGF$_P$	E-ICA-RGF$_C$
1	70.38	73.05	80.81	71.30	**95.81**	94.23
2	**100.00**	**100.00**	99.72	74.81	95.64	95.32
3	99.15	99.18	**99.76**	94.85	99.22	99.00
4	64.97	65.73	77.01	87.12	96.68	**97.78**
5	90.47	91.07	**95.36**	72.25	93.24	92.29
6	62.59	60.84	69.75	73.89	94.45	**97.57**
7	64.19	57.65	72.05	71.36	90.88	**92.08**
8	68.40	69.16	75.66	59.60	**92.76**	92.18
9	86.62	86.89	90.00	67.67	96.41	**97.91**
OA	69.18±2.74	68.59±2.64	76.61±2.00	73.24±1.88	94.54±1.83	**95.83±1.37**
AA	78.53±0.93	78.17±0.91	84.46±0.77	74.76±1.37	95.01±0.67	**95.37±0.62**

D. Comparisons with other-state-of-the art classifiers

Finally, we present comparisons of the proposed methods against the aforementioned state-of-the-art spectral-spatial classifiers, including GCK [25], RoF-MRF [27], RS-EMAPs [16], SVM-EPF [29] and IFRF [30], by considering different numbers of training samples (i.e., 10, 20, 30, 40, 50 samples per class). From Tables IX and X, we observe that the proposed methods outperform those methods in terms of classification accuracies and are more stable (with lower standard deviations) than other spectral-spatial classifiers.

TABLE VII: University of Pavia *ROSIS* image. Classification results achieved by the RoF classifier. For each method, "OA (%)", "AA (%)" and class-specific accuracies, "CA (%)" are reported.

Class	Original	E	E-ICA	E-RGF	E-ICA-RGF$_P$	E-ICA-RGF$_C$
1	77.46	81.53	84.53	82.64	**95.88**	88.64
2	**99.98**	99.93	99.88	75.94	95.09	95.07
3	99.18	99.33	**99.85**	94.63	99.49	98.21
4	83.31	77.94	83.09	92.54	95.77	**95.86**
5	93.12	92.92	**95.54**	81.97	92.55	90.39
6	67.94	59.69	67.62	83.50	95.04	**95.79**
7	69.48	64.09	64.28	82.72	**89.33**	87.86
8	74.21	74.31	76.84	57.16	89.58	**89.90**
9	86.00	86.01	84.78	82.09	**95.70**	94.97
OA	75.62±2.14	71.47±1.98	76.37±1.73	80.39±2.15	**94.06±1.02**	93.53±1.03
AA	83.41±0.89	81.75±0.92	84.05±0.85	81.47±0.95	**94.27±0.61**	92.97±0.63

TABLE VIII: Statistic of the McNemar's test .Each case of the Table represents Z_{cr} where c is the column and r is the row.

Methods	AVIRIS		ROSIS	
	RF	RoF	RF	RoF
Z_{cr}	E-ICA-RGF$_C$			
Original	38.51	25.34	50.34	40.87
E	39.27	24.83	51.24	42.65
E-ICA	26.14	29.53	37.25	37.23
E-RGF	16.15	18.21	39.41	30.24
E-ICA-RGF$_P$	1.21	-0.99	3.14	-2.01

TABLE IX: Indian Pines *AVIRIS* image. Classification accuracies obtained from the proposed methods in comparisons to other spatial-spectral classifiers.

Samples per class		RoF-MRF [27]	GCK [25]	RS-EMAPs [16]	SVM-EPF [29]	IFRF [30]	E-ICA-RGF$_P$	E-ICA-RGF$_C$
10 samples	OA	74.14±3.21	81.21±2.78	82.15±2.20	62.57±3.25	77.05±3.01	84.44±1.85	**86.30±2.15**
	AA	84.21±1.63	86.78±1.52	88.14±1.42	67.84±1.78	72.47±1.72	91.39±1.15	**91.86±1.39**
20 samples	OA	82.11±2.71	86.24±1.89	87.13±1.82	72.46±1.86	87.12±1.76	**90.92±1.63**	90.41±1.51
	AA	90.27±1.54	90.15±1.34	90.78±1.14	75.14±1.32	80.26±1.32	95.13±0.85	**95.78±0.81**
30 samples	OA	86.24±2.65	88.35±1.59	92.14±1.20	78.51±2.13	89.65±1.85	93.15±1.19	**93.43±0.99**
	AA	92.76±1.21	92.62±0.73	94.71±0.68	80.43±1.14	82.53±1.09	**96.23±0.59**	96.23±0.60
40 samples	OA	87.78±1.72	90.42±1.23	93.18±1.01	82.86±1.86	91.56±1.47	**95.35±0.92**	95.31±0.87
	AA	93.52±1.01	94.63±0.72	95.18±0.69	85.13±1.17	86.78±0.83	**97.53±0.38**	97.11±0.35
50 samples	OA	90.74±1.34	92.88±1.02	93.84±0.86	85.19±1.32	94.11±1.04	95.78±0.69	**95.90±0.72**
	AA	94.02±0.86	95.52±0.68	95.80±0.52	86.74±0.59	88.26±0.82	97.63±0.46	**97.89±0.48**

TABLE X: University of Pavia *ROSIS* image. Classification accuracies obtained from the proposed methods in comparisons to other spatial-spectral classifiers.

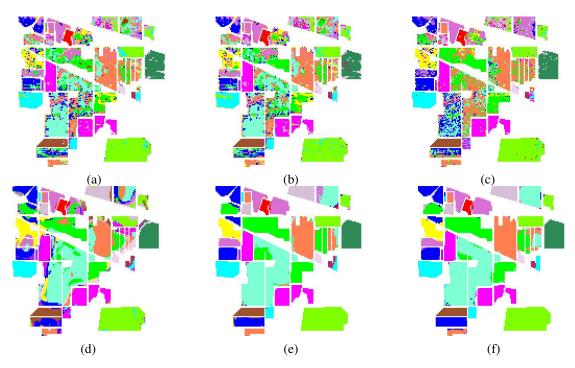| Samples per class		RoF-MRF [27]	GCK [25]	RS-EMAPs [16]	SVM-EPF [29]	IFRF [30]	E-ICA-RGF$_P$	E-ICA-RGF$_C$
10 samples	OA	79.14±3.65	84.53±3.19	85.16±3.23	76.35±4.08	77.18±3.31	**86.12±2.95**	85.79±3.15
	AA	85.62±1.23	89.79±1.14	89.15±1.27	78.42±1.49	78.51±1.24	90.46±1.12	**90.60±1.09**
20 samples	OA	84.88±2.72	89.53±2.81	89.61±2.28	86.21±2.89	86.78±2.34	92.56±1.96	**92.91±2.06**
	AA	89.13±1.08	92.28±1.14	92.87±1.01	87.14±1.21	83.52±1.19	94.02±0.76	**94.08±0.72**
30 samples	OA	89.27±2.35	90.62±2.59	92.87±1.98	88.19±2.46	89.32±2.11	94.54±1.83	**95.83±1.37**
	AA	91.46±1.27	94.53±0.97	94.46±0.87	87.51±1.31	84.41±1.18	95.01±0.67	**95.37±0.62**
40 samples	OA	91.30±1.31	93.11±1.28	93.16±0.99	91.75±1.11	92.07±1.16	**95.94±0.94**	95.86±0.86
	AA	92.98±0.85	95.12±0.96	94.98±0.81	93.78±0.96	92.22±0.83	95.73±0.73	**95.79±0.71**
50 samples	OA	92.75±1.00	94.25±1.04	94.18±0.91	93.05±0.98	93.18±0.89	**96.73±0.72**	96.49±0.70
	AA	94.11±0.65	96.03±0.45	95.87±0.53	93.78±0.61	92.86±0.86	**96.82±0.34**	96.24±0.36

Fig. 4: Indian Pines *AVIRIS* image. Classification results of (a) Original, OA = 62.04%. (b) E, OA = 62.84%. (c) E-ICA, OA = 69.37%. (d) E-RGF, OA = 73.27%. (e) E-ICA-RGF$_P$, OA = 94.87%. (f) E-ICA-RGF$_C$, OA = 95.32%.

E. Discussion

For spectral information based classification, the proposed E-ICA is superior to ICA applied to either high or low dimensional space. The rationale behind the success is that random subspace ensemble is introduced to combine the results obtained for the subsets via ICA techniques. In this case, we can enhance the classification performance without losing spectral information. The results in Tables II and III show that EBM and ERBM provide better classification results than JADE and FastICA in all the cases. The main reason is that EBM is a flexible density model to estimate the ICs [31], and ERBM, in addition to the use of this density model, incorporates sample dependence into the estimation procedure [32], [33].

The proposed spectral-spatial methods, E-ICA-RGF$_P$ and E-ICA-RGF$_C$, significantly improve the classification results compared with the use of only spectral information (E-ICA), indicating the importance of RGF in extracting spatial contextual information. The results shown in Tables IV, V, VI and VII indicate that both E-ICA-RGF$_P$ and E-ICA-RGF$_C$ are superior to E-ICA and E-RGF, demonstrating the effectiveness of combining ICA and RGF via an ensemble strategy. It should be noted that the orders of ICA and RGF are very important in our proposed methods. We have done an experiment by switching the ICA and RGF steps. Experimental results indicated that the classification accuracies are lower than our proposed methods, however, still higher than the ones of E-ICA and E-RGF. This is due to the fact that useful spectral information is lost after application of RGF, which prevents ICA to extract effective ICs. Moreover, other EPF techniques (e.g., relative total variation [44]) have also been tested
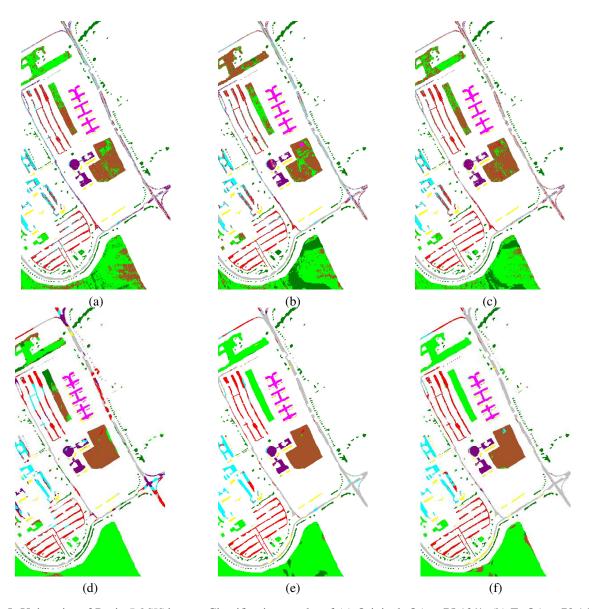
Fig. 5: University of Pavia *ROSIS* image. Classification results of (a) Original, OA = 75.13%. (b) E, OA = 70.14%. (c) E-ICA, OA = 78.06%. (d) E-RGF, OA = 79.47%. (e) E-ICA-RGF$_P$, OA = 94.93%. (f) E-ICA-RGF$_C$, OA = 93.87%.

and classification results were very close to the one obtained with the RGF.

An additional advantage of the proposed methods is the fact that parameter choice is not critical and does not need to be tuned finely. The users might select a small value of σ_r as in our case to increase the classification performance of the proposed method.

VII. CONCLUSION

In this paper, a new methodology for spectral-spatial supervised classification of hyperspectral remote sensing image has been proposed. The presented methodology uses a combination of ICA and edge preserving filter.
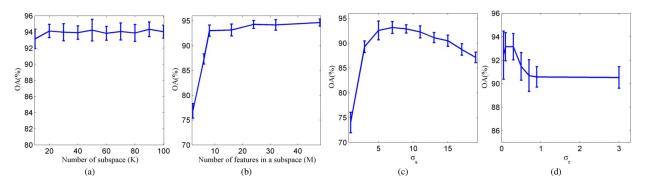
Fig. 6: Indian Pines AVIRIS image. Sensitivity of the proposed method: (a) influence of K ($M = 16$, $\sigma_s = 7$ and $\sigma_r = 0.1$), (b) influence of M ($K = 10$, $\sigma_s = 7$ and $\sigma_r = 0.1$), (c) influence of σ_s ($K = 10$, $M = 16$ and $\sigma_r = 0.1$) and (d) influence of σ_r ($K = 10$, $M = 16$ and $\sigma_s = 7$)
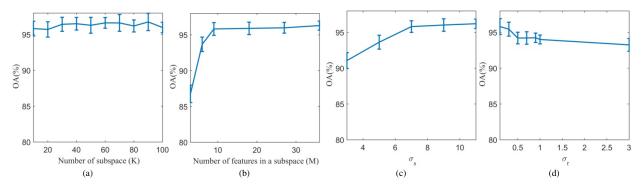

Fig. 7: University of *Pavia* image. Sensitivity of the proposed method: (a) influence of K ($M = 9$, $\sigma_s = 7$ and $\sigma_r = 0.1$), (b) influence of M ($K = 10$, $\sigma_s = 7$ and $\sigma_r = 0.1$), (c) influence of σ_s ($K = 10$, $M = 9$ and $\sigma_r = 0.1$) and (d) influence of σ_r ($K = 10$, $M = 9$ and $\sigma_s = 7$).

In particular, an ensemble strategy that combines the results obtained for ICA on the subsets is designed to retrieve effective spectral features in a high-dimensionality scenario without losing spectral information, i.e., where no prior dimensionality reduction is applied. More specifically, the ICA techniques are applied to each subset, which are randomly selected from the original feature space, and then the classification results obtained from ICs in each subset via a supervised classifier (e.g., RF or RoF) are integrated. Moreover, EBM and ERBM are introduced to the remote sensing community. EBM make uses of the non-Gaussianity property based on a flexible density model [31] and ERBM [32], [33] adds the use of sample dependence into the estimation procedure. Hence, both provide attractive alternatives for performing ICA for remote sensing applications as well.

The analysis is also extended to the spatial information domain with the definition of RGF for extracting spatial information. RGF effectively removes noise and small details while preserving large-scale structures. In this case, it can provide complementary spatial information of the structures present in the scene following the application of ICA in each subset. Moreover, the parallel (E-ICA-RGF$_P$) and concatenated (E-ICA-RGF$_C$) strategies are proposed to combine multiple features via ICA and RGF. The proposed methods (i.e., E-ICA-RGF$_P$ and E-ICA-RGF$_C$) are tested on two real hyperspectral images, which are different in terms of spectral/spatial resolution. The results show the effectiveness of the proposed methods in extracting spectral and spatial features, providing higher classification

accuracies when compared with state-of-the-art methods.

In our future studies, we will utilize other data sources (e.g., multi-temporal datasets) in the classification process. Furthermore, the optimization of our algorithms to reduce the computational burden and classification problems will be investigated.

ACKNOWLEDGMENT

REFERENCES

[1] C. I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. Plenum Publishing Co., 2003.

[2] C. I. Chang, *Hyperspectral Data Exploitation: Theory and Applications*. Wiley-Interscience, Hoboken, NJ, 2007.

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

[4] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, 2008.

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

[6] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832–844, August 1998.

[7] L. I. Kuncheva, J. J. Rodríguez, C. O. Plumpton, D. E. Linden, and S. J. Johnston, "Random subspace ensembles for fMRI classification," *IEEE Trans. Med. Imaging*, vol. 29, no. 2, pp. 531–542, 2010.

[8] V. N. Vapnik, *The Nature of Statistical Learning Theory*. Springer, 1995.

[9] 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, 2004.

[10] 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, 2005.

[11] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[12] J. C. W. Chan and D. Paelinckx, "Evaluation of Random Forest and AdaBoost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery," *Remote Sens. Environ.*, vol. 112, no. 6, pp. 2999–3011, Jun. 2008.

[13] J. J. Rodriguez, L. I. Kuncheva, and C. J. Alonso, "Rotation forest: A new classifier ensemble method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1619–1630, Oct. 2006.

[14] J. Xia, P. Du, X. He, and J. Chanussot, "Hyperspectral remote sensing image classification based on rotation forest," *IEEE Geosci. Remote Sensing Lett.*, vol. 11, no. 1, pp. 239 – 243, 2014.

[15] L. Zhang and P. N. Suganthan, "Random forests with ensemble of feature spaces," *Pattern Recognition*, vol. 47, no. 10, pp. 3429 – 3437, 2014.

[16] J. Xia, M. Dalla Mura, J. Chanussot, P. Du, and X. He, "Random subspace ensembles for hyperspectral image classification with extended morphological attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 4768–4786, 2015.

[17] J. Richards and X. Jia, *Remote sensing digitial image analysis*. 4th edition. Springer-Verlag Berlin Heidelberg, 2006.

[18] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Academic Press. Elsevier., 2010.

[19] J. Wang and C. I. Chang, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1586–1600, 2006.

[20] M. Dalla Mura, A. Villa, J. Benediktsson, J. Chanussot, and L. Bruzzone, "Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis," *IEEE Geosci. Remote Sensing Lett.*, vol. 8, no. 3, pp. 542–546, 2011.

[21] N. Falco, J. Benediktsson, and L. Bruzzone, "A study on the effectiveness of different independent component analysis algorithms for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2183–2199, June 2014.

[22] 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, 2005, pp. 1–4.

[23] C. Jutten, S. Moussaoui, and F. Schmidt, "How to apply ICA on actual data ? example of mars hyperspectral image analysis," in *Digital Signal Processing, 2007 15th International Conference on*, July 2007, pp. 3–12.

[24] Q. Du, I. Kopriva, and H. Szu, "Independent-component analysis for hyperspectral remote sensing imagery classification," *Optical Engineering*, vol. 45, no. 1, pp. 017 008–1017 008–13, 2006.

[25] J. Li, P. Reddy Marpu, A. Plaza, J. M. Bioucas-Dias, and J. A. Benediktsson, "Generalized composite kernel framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4816–4829, 2013.

[26] N. Falco, J. A. Benediktsson, and L. Bruzzone, "Spectral and Spatial Classification of Hyperspectral Images Based on ICA and Reduced Morphological Attribute Profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 6223–6240, nov 2015.

[27] J. Xia, J. Chanussot, P. Du, and X. He, "Spectral-spatial classification for hyperspectral data using rotation forests with local feature extraction and Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2532–2546, 2015.

[28] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. on Pattern Anal. Machine Intell.*, vol. 35, no. 6, pp. 1397–1409, June 2013.

[29] X. Kang, S. Li, and J. A. Benediktsson, "Spectral-spatial hyperspectral image classification with edge-preserving filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2666–2677, 2014.

[30] X. Kang, S. Li, and J. A. Benediktsson, "Feature extraction of hyperspectral images with image fusion and recursive filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3742–3752, 2014.

[31] X. Li and T. Adali, "Independent component analysis by entropy bound minimization," *IEEE Trans. Signal Processing*, vol. 58, no. 10, pp. 5151–5164, Oct 2010.

[32] X. Li and T. Adali, "Blind spatiotemporal separation of second and/or higher-order correlated sources by entropy rate minimization," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, March 2010, pp. 1934–1937.

[33] G.-S. Fu, R. Phlypo, M. Anderson, X.-L. Li, and T. Adali, "Blind source separation by entropy rate minimization," *IEEE Trans. Signal Process.*, vol. 62, no. 16, pp. 4245–4255, Aug 2014.

[34] J. Xia, L. Bombrun, T. Adali, Y. Berthoumieu, and C. Germain, "Classification of hyperspectral data with ensemble of subspace ICA and edge-preserving filtering," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, accepted for publication.

[35] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. on Neural Networks*, vol. 10, no. 3, pp. 626–634, May 1999.

[36] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *IEE Proceedings F (Radar and Signal Processing)*, vol. 140, pp. 362–370, December 1993.

[37] T. Adali, M. Anderson, and G. Fu, "Diversity in independent component and vector analyses: identifiability, algorithms, and applications in medical imaging," *IEEE Signal Proc. Mag.*, vol. 31, no. 3, pp. 18–33, 2014.

[38] T. Adali, H. Li, M. Novey, and J.-F. Cardoso, "Complex ICA using nonlinear functions," *IEEE Trans. Signal Processing*, vol. 56, no. 9, pp. 4536–4544, 2008.

[39] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proceedings of the Sixth International Conference on Computer Vision (ICCV)*, Washington, DC, USA, 1998, pp. 839–846.

[40] Q. Zhang, X. Shen, L. Xu, and J. Jia, "Rolling guidance filter," in *Computer Vision ECCV 2014*, ser. Lecture Notes in Computer Science, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Springer International Publishing, 2014, vol. 8691, pp. 815–830.

[41] J. Xia, J. Chanussot, P. Du, and X. He, "Rotation-Based Ensemble Classifiers for High-Dimensional Data," in *Fusion in Computer Vision*, B. Ionescu, J. Benois-Pineau, T. Piatrik, and G. Quénot, Eds. Springer, 2014, pp. 135–160.

[42] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri,

M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, no. S1, pp. S110–S122, Sep. 2009.

[43] M. Robnik-Sikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Mach. Learn.*, vol. 53, no. 1-2, pp. 23–69, 2003.

[44] L. Xu, Q. Yan, Y. Xia, and J. Jia, "Structure extraction from texture via relative total variation," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 139:1–139:10, Nov. 2012.