Automatic Framework for Spectral–Spatial Classification Based on Supervised Feature Extraction and Morphological Attribute Profiles

Pedram Ghamisi, *Student Member*, *IEEE*, Jón Atli Benediktsson, *Fellow*, *IEEE*, Gabriele Cavallaro, and Antonio Plaza, *Senior Member*, *IEEE*

***Abstract*—Supervised classification plays a key role in terms of accurate analysis of hyperspectral images. Many applications can greatly benefit from the wealth of spectral and spatial information provided by these kind of data, including land-use and land-cover mapping. Conventional classifiers treat hyperspectral images as a list of spectral measurements and do not consider spatial dependencies of the adjacent pixels. To overcome these limitations, classifiers need to use both spectral and spatial information. In this paper, a framework for automatic spectral–spatial classification of hyperspectral images is proposed. In order to extract the spatial information, Extended Multi-Attribute Profiles (EMAPs) are taken into account. In addition, in order to reduce the redundancy of features and address the so-called curse of dimensionality, different supervised feature extraction (FE) techniques are considered. The final classification map is provided by using a random forest classifier. The proposed automatic framework is tested on two widely used hyperspectral data sets; Pavia University and Indian Pines. Experimental results confirm that the proposed framework automatically provides accurate classification maps in acceptable CPU processing times.**

***Index Terms*—Extended Multi-Attribute Profile (EMAP), hyperspectral image analysis, random forest classification, supervised feature extraction (FE).**

I. Introduction

HYPERSPECTRAL imaging instruments are now able to capture hundreds of spectral channels from the same area on the surface of the Earth. By providing very fine spectral resolution with hundreds of (narrow) bands, accurate discrimination of different materials is possible. In parallel, due to recent advances in hyperspectral technology, the spatial resolution of the sensors is also becoming finer, which allows for a detailed characterization of spatial structures in the scene.

Supervised classification techniques play a key role in the analysis of hyperspectral images, and a wide variety of applications can be handled by successful classifiers in the literature [1], including: land-use and land-cover mapping, crop monitoring, forest applications, urban development, mapping, tracking, and risk management.

In the spectral domain, each spectral channel is considered as one dimension. By increasing the dimensionality in the spectral domain, theoretical and practical problems arise. Some of these problems are related to the curse of dimensionality, which is related to the unbalance between the (high) dimensionality of the input data and the (often limited) number of training samples used in the supervised classification process [2]. In [3], Landgrebe shows that too many spectral bands can be undesirable from the standpoint of expected classification accuracy. In other words, when the number of spectral bands (dimensionality) increases, with a constant number of training samples, the accuracy of the statistics estimation decreases. The aforementioned issue demonstrates that there is an optimal number of bands and that (given an available set of training samples) more features do not necessarily lead to better results. Therefore, feature reduction techniques may lead to better classification accuracies [4].

Conventional classifiers treat hyperspectral images as a list of spectral measurements with no particular arrangement [5] and do not consider spatial dependencies of adjacent pixels. In other words, conventional techniques classify images only based on their spectral information alone. Therefore, these approaches discard information associated with the spatial correlations among distinct pixels in the image. In order to address the aforementioned issue, the consideration of both spectral and spatial information has been widely explored in the literature [6]. In addition, spatial information can provide additional information related to the shape and size of different structures, which generally leads to better classification accuracies and classification maps.

Two strategies are commonly used in order to characterize spatial information: crisp neighborhood system and adaptive neighborhood system. Although the first one mostly considers spatial and contextual dependencies in a predefined

Manuscript received September 19, 2013; revised December 13, 2013; accepted January 06, 2014. Date of publication January 29, 2014; date of current version August 01, 2014. This work was supported in part by the Icelandic Research Fund for Graduate Students.

P. Ghamisi and G. Cavallaro are with the Faculty of Electrical and Computer Engineering, University of Iceland, Reykjavik 110, Iceland (e-mail: p.ghamisi@gmail.com; cavallaro.gabriele@gmail.com).

J. A. Benediktsson is with the Engineering Research Institute, University of Iceland, Reykjavik 107, Iceland (e-mail: benedikt@hi.is).

A. Plaza is with the Technology of Computers and Communications, Escuela Politecnica de Caceres, University of Extremadura, E-10071 Caceres, Spain (e-mail: aplaza@unex.es).

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/JSTARS.2014.2298876

neighborhood system, the latter is more flexible and it is not confined to a given neighborhood system. One way for extracting spatial information with crisp neighborhood is to consider Markov random field (MRF) modeling. MRF is a family of probabilistic models that can be described as a 2-D stochastic process over discrete pixel lattices [7]. There is extensive literature on the use of MRFs in classification, such as [8], [9]. However, the main disadvantages of considering a set of crisp neighbors are that 1) the standard neighborhood system may not contain enough samples, which decreases the effectiveness of the classifier (in particular, when the input data set is of high resolution and the neighboring pixels are highly correlated [6]) and 2) a larger neighborhood system leads to computationally intractable problems [6]. In order to address the aforementioned issues, adaptive neighborhood systems can be taken into consideration. A possible way to develop adaptive neighborhood systems is to use different types of segmentation methods. Image segmentation is a procedure which can be used to modify the accuracy of classification maps [10]. To make such an approach effective, an accurate segmentation of the image is needed [11]. Several works have previously explored the extraction of spatial information using segmentation techniques (e.g., [12]–[14]). Another set of methods which can extract spatial information by using adaptive neighborhood systems relies on morphological filters. Pesaresi and Benediktsson [15] used morphological transformations to build a so-called Morphological Profile (MP). In [16], the MP was used to handle hyperspectral images and named Extended Morphological Profile (EMP) in this context. Attribute Profiles (APs) constitute another extension of the concept of MP and provide a multilevel characterization of an image by the sequential application of morphological attribute filters, which model different specifications of the structural information contained in the scene [17]. APs provide a powerful tool to increase the discrimination of different classes [17], [18]. However, there are two main difficulties associated with the concept of Extended Multi-Attribute Profiles (EMAP), including: 1) how to establish which attributes lead to a better discrimination for different classes and 2) how to determine which values should be considered in order to initialize each AP.

In this paper, a new fully automatic approach is proposed for accurate classification of hyperspectral images. Although the presented framework can also be used for classification of multispectral images with coarser spectral resolution, it is used here for spectral–spatial classification of hyperspectral images. In order to extract the spatial information, EMAP [17] are automatically generated by the proposed framework. In order to reduce the redundancy of the data and address the so-called curse of dimensionality, different supervised feature extraction (FE) techniques are also included in the proposed framework. The final classification map is provided by a Random Forest (RF) classifier [19], [20]. In order to handle high-dimensional data, RF and SVM have been widely considered as the most powerful classifiers since they are robust when handling high-dimensional data with a limited number of training samples. Both the SVM and RF classifiers are comparable in terms of classification accuracies and have been widely used for the purpose of hyperspectral image classification. However, while both methods are shown to be effective classifiers for nonlinear classification problems, SVM requires a computationally demanding parameter tuning process in order to achieve optimal results, whereas RF does not require such a tuning process. In this sense, RF is faster than SVM. In this paper, our main objective is obtaining good classification accuracies in an acceptable CPU processing time. In addition, several studies such as [21] have reported that Hughes phenomenon [2] is more evident when the number of dimensions is high and the data are classified by SVM instead of RF. The proposed approach is tested using two well-known data sets collected by the reflective optics spectrographic imaging system (ROSIS) over the city of Pavia, Italy, and by the Airborne Visible Infra-Red Imaging Spectrometer (AVIRIS) over the Indian Pines region in northwestern Indiana. The experimental results confirm that the presented framework is able to classify hyperspectral images efficiently both in terms of classification accuracies and CPU processing time. It should be noted that the proposed approach is fully automatic and there is no need to initialize any parameters empirically. The main contribution of this paper compared to other works on EMAP is that, in most of previous works, the EMAP is built using an unsupervised FE approach such as principal component analysis (PCA), independent component analysis (ICA), and Kernel PCA, while in this work, we explore the use of supervised FE for this purpose. Another main difference is concerned with the automatic nature of our approach in the sense that, in previous works, threshold values for making EMAP needed to be initialized manually, while, in the present framework, a general range of parameter values is used to make the parameter selection automatical. Another important difference is that, in most previous works, the outcome of the AP is directly used for classification, while in our framework, we use a second FE strategy prior to classification. As shown by the present contribution, the results of the second FE step are used for classification purposes by our proposed framework, with concatenating features of both FE steps in one vector and then performing the classification step.

This remainder of the paper is organized as follows: the proposed framework is discussed in Section II. Section III is devoted to validating the framework via extensive experimental results. Section IV outlines the main conclusions and provides hints at plausible future research lines.

II. Framework

In the proposed framework, supervised FE is first performed on the input data and the first features with cumulative eigenvalues above 99% are retained. In the case of discriminant analysis FE (DAFE), the criterion is related to the size of the eigenvalues of the scatter matrices. In the case of decision boundry FE (DBFE), it is related to the size of the eigenvalues of the decision boundary feature matrix (DBFM). Let us consider ζ_φ as the output of this step. Then, EMAPs are built on the first few features and the resulting features are concatenated into one stacked vector. In order to reduce the redundancy of the stacked vector, a supervised FE step is performed once again. Let ζ_ω be the output of this step. The final classification map is provided by performing RF classification on the stacked vector, $\zeta = [\zeta_\varphi, \zeta_\omega]^T$. Fig. 1 illustrates the proposed framework by a

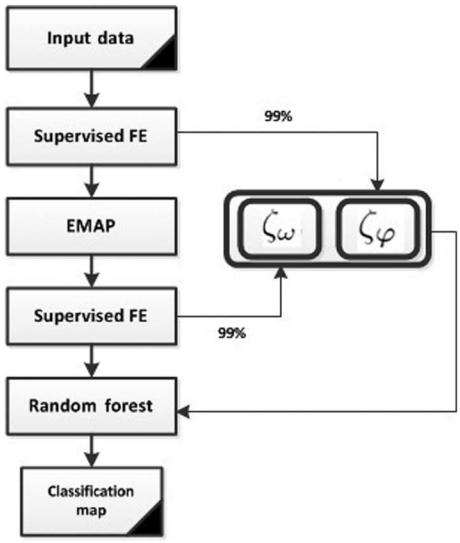

Fig. 1. Flowchart of the proposed framework.

flowchart. In the following, the individual parts of the proposed framework will be discussed in detail.

A. Feature Extraction (FE)

FE consists of finding a set of vectors that represent an observation while reducing its dimensionality. FE techniques can be grouped into two categories: unsupervised approaches and supervised approaches, where the former is used for the purpose of data representation and the latter is considered for overcoming the Hughes phenomena and reducing the redundancy of data in order to improve classification accuracies. PCA is an example of unsupervised FE. PCA does not find optimum feature sets in the sense of class discrimination and discards class specific information. Therefore, for image classification, supervised FE may lead to higher classification accuracies. From one point of view, supervised FE techniques can be split into two categories: parametric and nonparametric. The main disadvantages of nonparametric FE techniques are that they do not have an assumption on the underlying density functions in the data. Therefore, FE for nonparametric classifiers is often not feasible or very time consuming. On the contrary, although the computation cost of nonparametric classifiers is often much larger than that of parametric classifiers, there are some cases where the use of nonparametric FE is desirable. For example, if the underlying densities are unknown or problems involve complex densities, which cannot be approximated by a common parametric density functions, the use of a nonparametric classifier is important [22]. In this work, two approaches are considered for supervised FE: DAFE and DBFE. Below, the considered supervised FE techniques are briefly explained.

1) Discriminant Analysis FE (DAFE): This approach is widely used for dimension reduction in classification problems [23]. Since, DAFE uses the mean vector and the covariance matrix of each class, it is considered as supervised FE. In DAFE, within-class, between-class, and mixture scatter matrices are usually considered as the criteria for class separability. DAFE is fast and works well when the distribution of the data is normal. Otherwise, the performance of DAFE may not be satisfactory. Another problem associated with this method is that if the difference in the class-mean vectors is small, the feature chosen may not be reliable. Similarly, if one class-mean vector is very different from others, its class will dominate the others in the computation of the between-class covariance matrix [24]. As a consequence, the FE process may be ineffective. In addition, DAFE performs the computations with full dimensionality, which requires a large number of training samples in order to accurately estimate parameters. A main shortcoming of DAFE is that DAFE is not full rank and its rank is at maximum $L-1$ where L is the number of classes. Let us assume that the rank of the within-class scatter matrix is u, in this case, only $\min(L-1,u)$ features are selected by using DAFE. Since the complexity of the data in real scenarios could be quite high, using only $L-1$ features may not be enough to fully characterize the data.

2) Decision Boundary Feature Matrix (DBFE): This method was proposed in [25] where it was shown that both discriminantly informative and redundant features can be extracted from the decision boundary between two classes. The features are extracted from the DBFM. In order to obtain the same classification accuracy as in the original space, keeping the eigenvectors of the DBFM corresponding to nonzero eigenvalues is crucial. The performance of this method does not deteriorate even when there is no difference in the mean vectors or covariance matrices. It should be noticed that this approach does not rely on the number of classes in the same way as DAFE. The efficiency of DBFE is highly dependent on the quality and number of training samples, which is not desirable. Another shortcoming of DBFE is that it can be computationally intensive.

B. Extended Multi-Attribute Profile (EMAP)

Mathematical morphology [26]–[29] is a well-established framework which provides operators able to high-quality spatial features. Fundamental mathematical morphology operators, such as Erosion and Dilation (and their combinations: opening and closing), examine the geometrical structures in the image by matching them to small patterns called structuring elements. Depending on the shape and size of the structuring element, undesirable effects can occur in the filtered image; in particular, geometrical characteristics of the structures can be distorted or completely lost. In this work, morphological operators are considered which perform transformations by reconstruction, a class of connected filters [30]. Specifically, they act on connected components, i.e., flat regions of a gray scale image, which are either completely removed or preserved according to their interaction with the structuring element adopted by the transformation.

1) Attribute Filters Based on Tree Representation: Attribute filters [31] are flexible operators that can perform simplification of a grayscale image driven by an arbitrary measure which can be related to characteristics of regions in the scene such as the scale, shape, contrast, etc. Improvements in terms of capability in modeling the spatial information are achievable since these

operators are not based on fixed structuring elements, and the image transformation is only computed by merging its connected components. The idea is to extract different types of information, represented by the attributes, from different flat regions, i.e., parts of the scene with the same gray levels. Attribute filters are efficiently implemented with an equivalent representation of the image as a tree [32].

In particular, a thresholding operation of all the mapped values present in the image f, results in upper and lower level sets which are connected components (i.e., flat zones) that can be grouped in the following sets:

$$\mathcal{U}(\mathrm{f}) = \{X : X \in \mathcal{CC}([f \geq \lambda]), \lambda \in \mathbb{Z}\}$$
$$\mathcal{L}(\mathrm{f}) = \{X : X \in \mathcal{CC}([f < \lambda]), \lambda \in \mathbb{Z}\}$$

where $\mathcal{CC}(f)$ being the connected components of the generic image f. There is an inclusion relationship [33] between the connected components extracted by both the upper or lower level sets [belonging to $\mathcal{U}(\mathrm{f})$ or $\mathcal{L}(\mathrm{f})$, respectively]. This property allows for the association of a node in the tree to each connected component and thus represent the image as a hierarchical structure: the max-tree and min-tree [32] structures represent, respectively, the components in $\mathcal{U}(\mathrm{f})$ and $\mathcal{L}(\mathrm{f})$ with their inclusion relations by the thresholding operations. Attribute filters are shape preserving, since they never introduce new edges in an image [32], and operate on regions according to the result of a binary predicate P. In particular, the filtering criteria usually determine whether the value of an attribute α of a given connected component CC verifies a predicate: $P = \alpha(CC) \geq \lambda$ with $\{\alpha(CC), \lambda\} \in \mathbb{R}$ or $\mathbb{Z}$, where λ is a threshold value. When attribute filters are applied to the tree representation of the image, the operator leads to a pruning of the tree by removing the nodes whose associated regions do not fulfill P. Two different filtering approaches have been proposed: pruning the tree by removing whole branches and pruning by not removing all the branches [34]. Attributes can be purely geometric (e.g., area, length of the perimeter, and moment of inertia) or textural (e.g., standard deviation and entropy). A very detailed characterization of features is usually obtained.

2) Attribute Profiles: The spatial features can be derived in different ways such as with Gray-Level Co-occurrence Matrix (GLCM), Differential Morphology Profiles (DMPs), or Urban Complexity Index (UCI) [35]. Here, we propose to use EMAPs instead of GLCM, DMPs, and UCI. The use of EMAPs based on mathematical morphology concepts exhibits some desirable features in the context of hyperspectral image classification. Specifically, they offer a very flexible approach since they can perform the processing based on many different types of attributes. In fact, the attributes can be of any type. For example, they can be purely geometric, or related to the spectral values of the pixels, or on different characteristics. Furthermore, an efficient implementation based on tree representation has been used. In summary, EMAP offers a different strategy to include spatial information when compared to GLCM or UCI.

The spatial information belonging to different features present in very high-resolution data can be efficiently exploited by considering a multilevel approach based on morphological attribute filters. In particular, APs define a general set of profiles which take advantage of the flexibility of the attribute filters in order to better investigate the scene. According to the type of the criteria (increasing, nonincreasing), APs are defined differently. In the case of increasing attributes, the AP is a sequence of attribute openings and closings which include morphological opening and closing profiles by reconstruction [36]. On the other hand, when dealing with increasingness criteria, attribute thinning and thickening over a multilevel approach is applied. The result is the obtained attribute thinning and thickening profiles, which perform a multilevel analysis of the image based on attributes (represented by ordered criteria) not necessary related to the scale of the structures of the image.

APs can be, therefore, regarded as more effective filters than MPs, this is because the latter perform a partial characterization of the objects in the scene as a consequence of the fact that structuring elements are intrinsically unsuitable to describe features related to the graylevel characteristic of the region. Another considerable advantage is that APs are computed according to an effective implementation based on max-tree and min-tree representations, which lead to a reduction of the computational load when compared with conventional profiles built with operators by reconstruction.

3) Extended Attribute Profiles: Since hyperspectral sensors collect information in several spectral bands, Extended Attribute Profiles (EAPs) which are based on morphological attribute filters are adopted in order to perform the analysis of hyperspectral high-resolution images. The extension to multivalued images is not a trivial task, and morphological operators compute their function in a different domain which becomes a subset of the multivariate domain, where the ordering of the mapped vector values is not defined anymore. The EAPs rely on the application of the APs to hyperspectral data and they are simply defined as [36]

$$EAP = \{AP(PC_1), AP(PC_2), \ldots, AP(PC_c)\} \quad (1)$$

where PC name denotes a principal component obtained after applying PCA [37]. As mentioned before, PCA does not find optimum feature sets in the sense of class discrimination and discards class specific information. Therefore, for image classification, supervised FE leads to higher classification accuracies since such approaches provide optimal features with respect to class specific information. The EAP includes in its definition the EMP since the operators by reconstruction can be viewed as a particular set of morphological attribute. Since the modeling of spatial features is performed by attribute filters, this approach leads to a great flexibility, and the computation of the filters on the max-tree structure reduces the computational complexity with respect to EMP since the tree is built once for each principal component and filtered multiple times, according to the required number of levels.

4) Extended Multi-Attribute Profiles: APs extract efficiently spatial features by considering different attributes; for this reason, EMAPs merge different EAPs in a single data structure [36]

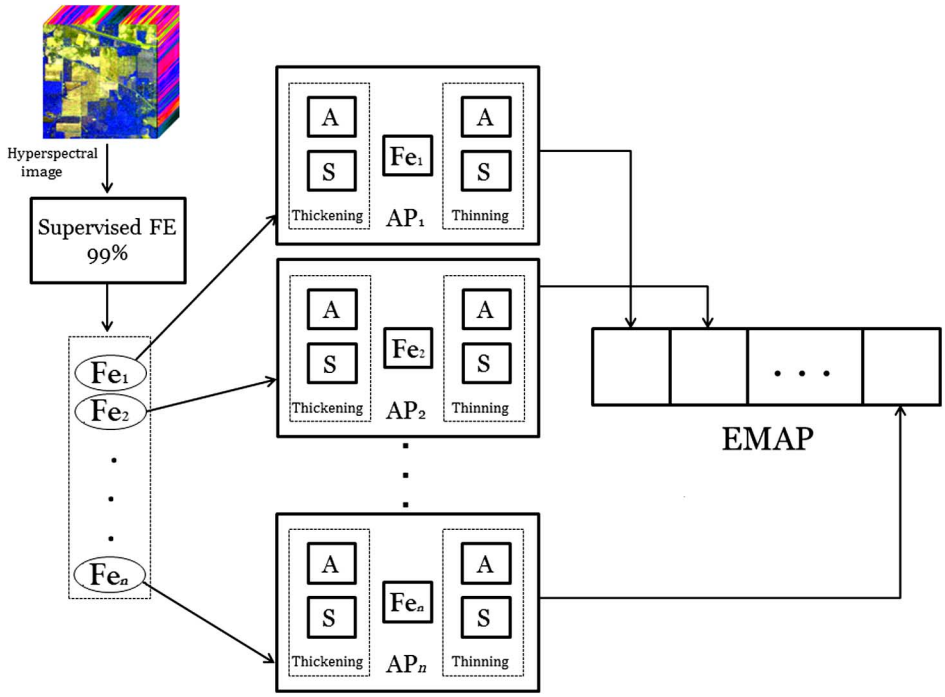

Fig. 2. Automatic framework for the construction of EMAPs. First, a supervised FE step is performed on the input data and the first features with cumulative eigenvalues above 99% are kept. Then, EMAPs are built for the first few features and the output features are concatenated into one stacked vector.

$$EMAP = \{EAP_{a_1}, EAP'_{a_2}, \ldots, EAP'_{a_m})\}. \quad (2)$$

Since the dimensionality of the features is increased, the EMAP has much greater capabilities in extracting spatial information than a single EAP but, at the same time, the computational cost of processing these features is slightly higher since the max-tree and min-tree are computed only once for each PC and they are filtered with different attributes at different levels.

5) Automatic Framework: Now, an automatic framework is introduced in order to solve issues such as the automatic selection of the attributes that lead to a best possible discrimination between the classes or the automatic identification of the most appropriate values to initialize each AP. Fig. 2 shows the general idea of the automatic framework for the construction of EMAPs. Although the APs can be constructed by using a wide variety of attributes, in the automatic framework, only the area and standard deviation attributes are used, since the aforementioned attributes can be adjusted in an automatic way and are well related to the object hierarchy in the images. The standard deviation is adjusted with respect to the mean of the individual features, since the standard deviation shows dispersion from the mean [21]. Therefore, λ_s is initialized so as to cover a reasonable amount of deviation in the individual feature, which is mathematically given by

$$\lambda_s(Fe_i) = \frac{\mu_i}{100}\{\sigma_{min}, \sigma_{min} + \delta_s, \sigma_{min} + 2\delta_s, \ldots, \sigma_{max}\} \quad (3)$$

where Fe_i denotes the ith feature obtained by a supervised FE. μ_i is the mean of the ith feature and σ_{min}, σ_{max}, and δ_s are 2.5%, 27.5%, and 2.5%, respectively, which leads to 11 thinning and 11 thickening operations. It should be noticed that the above-mentioned parameters have been tested on other well-known data sets with different spatial resolutions in [18] and results confirm that these parameters are data set distribution independent and can provide excellent results in terms of classification accuracies.

With regard to the adjustment of λ_a for the area attribute, the resolution of the image should be taken into account in order to construct the EAP [18]. The automatic construction of the attribute area is accomplished by the following expression:

$$\lambda_a(Fe_i) = \frac{1000}{\upsilon}\{a_{min}, a_{min} + \delta_a, a_{min} + 2\delta_a, \ldots, a_{max}\} \quad (4)$$

where a_{min} and a_{max} are initialized by 1 and 14, respectively, with a stepsize increase of δ_a equal to 1. The EAP for the area attribute includes 14 thinning and 14 thickening operations for each feature. Each level is provided in square meters by considering the resolution of the image υ in meters. Each profile covers structures in the range of $1000-14\ 000\ \mathrm{m}^2$, which might be a reasonable range of sizes for different structures in both urban and rural cases in remote sensing images [18]. However, different ranges can be considered for different applications.

Regarding (3) and (4), the used parameters have been tested on other well-known data sets with different spatial resolutions in [18] and results confirm that these parameters are data set distribution independent and can provide excellent results in terms of classification accuracies. In other words, those parameters do not need to be tuned for different data sets with different spatial resolutions. In the introduced framework, one only needs to establish a range of parameter values in order to automatically obtain a classification result with high accuracy for different data sets. It turns out that the used parameter ranges

have been tested on other well-known data sets with different spatial resolutions, such as the ones described in [18], and the obtained results confirm that these parameters are data set independent. In other words, those parameter ranges can be fixed for different data sets with different spatial resolutions. In [38], it was shown that the automatic scheme with only two attributes (area and standard deviation) can provide results comparable with a manual scheme with four attributes in terms of classification accuracy and CPU processing time.

C. Fusion of Extracted Features Via Vector Stacking

As indicated in Fig. 1, the input data are transformed by a supervised FE and only the first few features are used in order to reduce redundancy in the data while keeping most of the data variance. Then, the EMAP is computed by using only the first effective features that correspond to 99% of the eigenvalues.

Let ζ_φ be the set of features retained. Then, MAP is performed on each feature of ζ_φ and the output features are concatenated into one stacked vector. In order to address the so-called curse of dimensionality and reduce the redundancy of the stacked vector, a supervised FE step is performed once again. Let ζ_ω be the output of this step consisting of the features with cumulative eigenvalues above 99%. The final classification map is achieved by performing RF classification on the stacked vector; $\zeta = [\zeta_\varphi, \zeta_\omega]^T$.

D. Random Forest (RF)

RF was first introduced in [19]. It is an ensemble method for classication and regression. Ensemble classifiers get their name from the fact that several classifiers are trained and their individual results are then combined through a voting process. For the classification of an object from an input vector, the input vector is run down each tree in the forest. Each tree provides a unit vote for a particular class and the forest chooses the classication having the most votes. Based on [20], the computational complexity of the RF algorithm is $cT\sqrt{MN}\log(N)$ where c is a constant, T denotes the number of trees in the forest, M is regarded as the number of variables, and N is the number of samples in the data set. It is easy to infer that RF is not computationally intensive but demands a considerable amount of memory, since it is necessary to store an $N \times T$ matrix in the process. RF has several advantages, such as the capacity to provide good classification accuracies and to handle many variables. Another advantage of the RF classifier is that it is insensitive to noise in the training samples. In addition, RF provides an unbiased estimate of the test set error as trees are added to the ensemble, with almost no sensitivity to overfitting issues.

III. Experimental Results

A. Data Description

Two hyperspectral data sets were used in experiments. They are described as follows.

1) Pavia University: The first test case is a hyperspectral data set captured on the city of Pavia, Italy by the Reflective Optics Spectrographic Imaging System (ROSIS-03) airborne instrument. The ROSIS-03 sensor has 115 data channels with a spectral coverage ranging from 0.43 to 0.86 μm. The data have been corrected atmospherically, but not geometrically. The spatial resolution is 1.3 m per pixel. The data set covers the Engineering School at the University of Pavia and consists of different classes including: trees, asphalt, bitumen, gravel, metal sheet, shadow, bricks, meadow, and soil. In our experiments, 12 noisy data channels were eliminated and 103 data channels used for processing. The original data set comprises 640×340 pixels. Fig. 3(a) shows a false color composite of Pavia University and Fig. 3(b) shows a fixed training set that will be used for training purposes in this paper. Fig. 3(c) shows the available reference data for the scene. The number of available test and training samples is listed in Table I.

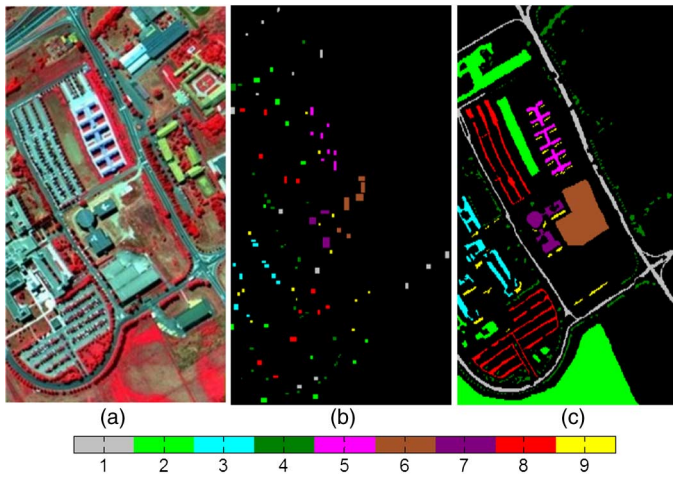

Fig. 3. The ROSIS-03 Pavia University data set: (a) false color image, (b) training samples, and (c) test samples, where each color represents a specific information class. The information classes are listed in Table I.

2) Indian Pines Data: The second data set used in experiments is the well-known data set captured on Indian Pines (NW Indiana) in 1992 comprising 16 classes (see Fig. 4), mostly related to different land covers. The data set consists of 145×145 pixels with spatial resolution of 20 m/pixel. In this work, 200 data channels are used, i.e., after the removal of the spectral bands affected by atmospheric absorption. The number of training and test samples is displayed in Table II.

It should be noted that, in addition to selecting widely used data sets in the hyperspectral imaging community, we have used exactly the same training and test samples that have been considered in most works related to spectral–spatial classification of hyperspectral images. Some of the works that have considered exactly the same training and test samples are those in [9], [39], and [40]. In other words, we not only used the same number of training and test samples adopted by other state-of-the-art methods, but also these samples have exactly the same spatial locations in the data. This way of using the training and test samples makes this work fully comparable with other spectral and spatial classification techniques reported in the literature. In order to keep consistency with previous results, each method was run only once since we have not used different training and test samples, but instead used exactly the same samples as adopted in the previous studies.

TABLE I
PAVIA UNIVERSITY: NUMBER OF TRAINING AND TEST SAMPLES ALONG WITH CLASSIFICATION ACCURACIES FOR THE RAW SPECTRAL DATA IN PERCENTAGE

Class Number	Class Name	Number of samples training	Number of samples test	Raw (103)
1	Asphalt	548	6631	80.8
2	Meadows	540	18 649	56.1
3	Gravel	392	2099	53.5
4	Trees	524	3064	98.7
5	Metal sheets	256	1345	99.1
6	Soil	532	5029	78.1
7	Bitumen	375	1330	84.3
8	Bricks	514	3682	91.0
9	Shadows	231	947	98.3
Kappa	–	–	–	0.6511
OA	–	–	–	71.64
AA	–	–	–	82.25

The number of features is given in the parentheses.

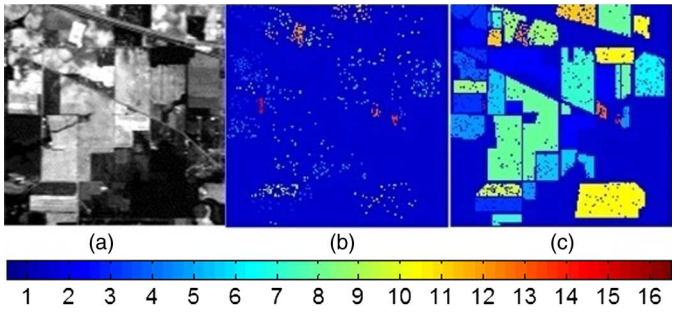

Fig. 4. AVIRIS Indian Pines data set: (a) spectral band number 27 ($\lambda = 646.72$ nm), (b) training samples, and (c) test samples, where each color represents a specific information class. The information classes are listed in Table II.

TABLE II
INDIAN PINES: NUMBER OF TRAINING AND TEST SAMPLES ALONG WITH CLASSIFICATION ACCURACIES FOR THE RAW SPECTRAL DATA IN PERCENTAGE

Class Number	Class Name	Number of samples training	Number of samples test	Raw (200)
1	Corn-notill	50	1384	57.5
2	Corn-mintill	50	784	58.6
3	Corn	50	184	85.8
4	Grass-pasture	50	447	85.6
5	Grass-trees	50	697	79.9
6	Hay-windrowed	50	439	94.7
7	Soybean-notill	50	918	78.5
8	Soybean-mintill	50	2418	58.8
9	Soybean-clean	50	564	62.9
10	Wheat	50	162	96.3
11	Woods	50	1244	88.5
12	Bldg-grass-tree-drives	50	330	57.5
13	Stone-steel-towers	50	45	93.3
14	Alfalfa	15	39	53.8
15	Grass-pasture-mowed	15	11	81.8
16	Oats	15	5	100
Kappa	–	–	–	0.6642
OA	–	–	–	70.24
AA	–	–	–	76.98

The number of features is given in the parentheses.

B. Experimental Setting

In experiments, the input image is transformed by a supervised FE and the first features with cumulative eigenvalues above 99% are retained, since they are expected to contain most of the variance in the original data sets. In the proposed framework, there is a second FE step which is conducted using the same criterion. For simplicity, the names of the different classifiers will be referred hereinafter as follows:

1) *Raw*: when the input data are classified by RF.
2) *Spec*: when only Spectral information resulting from the first supervised FE is classified by RF.
3) *AP*: when the selected features are used in order to produce the EMAP and classified by the RF.
4) n_m: when a supervised FE is performed for the second time and the output is classified by the RF. We have decided to use the name n_m in this case, where n FE approach means that the input data set first is transformed by n FE approach and the EMAP is then transformed by m FE approach. As an example, DB_DA means that the raw data were transformed by DBFE and the EMAP by DAFE.
5) ζ_{DA}: stacked vector consisting all features resulting from the first and second supervised FE. The suffix DA refers to the second FE technique.
6) ζ_{DB}: stacked vector consisting of all features resulting from the first and second supervised FE. The suffix DB refers to the second FE technique.

In the following, the number of features for Spec indicates the number of features with cumulative eigenvalues of more than 99% after performing DAFE or DBFE on the raw data. For example, it can be seen from Table III that six features are kept for Spec. It means that, first, the input data are transformed by DAFE and the first features with cumulative eigenvalues of more than 99% were kept (six features). These six features are used as a baseline for constructing the EMAP. Then, 14 thinning and 14 thickening are produced for the area attribute and 11 thinning and 11 thickening are produced for the standard deviation attribute. Therefore, each feature was used to produce 50 attributes and, by considering the feature itself in that vector, we have 51 features for each feature obtained by DAFE ($6 \times 51 = 306$ features for AP). Then, the second FE was performed and the first features with cumulative eigenvalues of more than 99% were kept. In this way, for Table III, DA_DA and DA_DB consist of 8 and 24 features, respectively. ζ_{DA} is the combination of Spec and DA_DA ($6 + 8 = 14$) and ζ_{DB} is the combination of Spec and DA_DB ($6 + 24 = 30$).

The way we calculate the CPU processing of each method is listed as follows:

1) *Spec*: CPU processing time of the first FE plus the CPU processing time of the corresponding RF classification.
2) *AP*: CPU processing time of the first FE plus the CPU processing time of producing EMAP plus the CPU processing time of the corresponding RF classification.
3) n_m: CPU processing time of the first FE plus the CPU processing time of producing EMAP plus the CPU processing time of the second FE plus the CPU processing time of the corresponding RF classification.
4) $\zeta_{\mathrm{DA\ or\ DB}}$: CPU processing time of the first FE plus the CPU processing time of producing EMAP plus the CPU

TABLE III
PAVIA UNIVERSITY: CLASSIFICATION ACCURACIES OBTAINED BY DIFFERENT METHODS FOR THE ROSIS PAVIA UNIVERSITY SCENE AFTER APPLYING DAFE

Class Number	Class Name	Spec (6)	AP (306)	DA_DA (8)	DA_DB (24)	ζ_{DA} (14)	ζ_{DB} (30)
1	Asphalt	82.9	98.0	98.2	97.1	**98.3**	97.0
2	Meadows	71.4	92.6	70.1	82.5	69.0	**96.5**
3	Gravel	69.5	81.0	**93.4**	89.7	91.6	86.4
4	Trees	92.2	97.8	**99.5**	99.3	**99.5**	**99.5**
5	Metal sheets	99.9	99.8	99.9	99.8	**100**	99.8
6	Soil	87.8	98.6	99.8	**99.9**	99.7	**99.9**
7	Bitumen	84.5	**100**	99.7	**100**	99.7	99.8
8	Bricks	85.1	96.1	**99.4**	99.3	**99.4**	**99.4**
9	Shadows	**97.7**	94.5	92.5	91.0	92.4	91.5
Kappa	–	0.7426	0.9317	0.8258	0.8862	0.8187	**0.9619**
OA	–	79.63	94.77	86.13	91.13	85.54	**97.11**
AA	–	85.71	95.41	94.75	95.42	94.43	**96.68**

The number of features used for classification purposes is reported in the parentheses.

TABLE IV
PAVIA UNIVERSITY: CPU PROCESSING TIME (IN SECONDS) OF DIFFERENT METHODS AFTER APPLYING DAFE

Spec	AP	DA_DA	DA_DB	ζ_{DA}	ζ_{DB}
7	131	51	161	48	165

processing time of the second FE plus the CPU processing time of corresponding the RF classification.

The following measures are used in order to evaluate the performance of different classification methods.

1) *Average accuracy (AA):* this metric shows the average value of the class classification accuracy.
2) *Overall accuracy (OA):* this metric refers to the number of samples, which are classified correctly divided by the number of test samples.
3) *Kappa coefficient:* this metric provides information regarding the agreement corrected by the level of agreement that could be expected due to chance alone.
4) *CPU processing time:* this metric shows the speed of different algorithms. It should be noted that, since in all algorithms (except Spectral), EMAP is carried out, the CPU processing time of this step is discarded from all methods. Hence, the CPU processing time is only provided for AP, $\mathrm{AP} + \mathrm{spectral}$, DAFE, and NWFE. All methods were implemented in MATLAB on a computer having Intel(R) Pentium(R) 4 CPU 3.20 GHz and 4 GB of memory.

C. Experimental Results

1) Pavia University: Table III gives information related to the classification accuracies of different methods after applying DAFE, with the corresponding CPU processing times listed in Table IV. As it can be observed from Tables I and III, the Spectral classification with only six features improves the OA of the Raw, data with 103 bands by 8%. Also, the class accuracies of Meadows and Gravel classes can be improved. Specifically, many samples of Meadows are misclassified as belonging to soil. Moreover, many samples of Gravel are misclassified as belonging to Asphalts and Bricks.

As can be seen from Table III, ζ_{DB} (consisting of 30 features) outperforms other methods significantly. ζ_{DB} improves the OA of Spectral, AP, $\mathrm{DA\_DA}$, $\mathrm{DA\_DB}$, and ζ_{DA} by 18%, 2.5%, 11, 6%, and 12%, respectively.

As it was already observed for the AVIRIS Indian Pines data set, AP achieves the best OA after ζ_{DB} since AP can model spatial dependencies of different objects by considering an adaptive neighborhood system. As can be seen from Table V, the OA of Spectral with 29 features improves the OA of Raw (with 103 bands) by 8%. Another observation is that AP, $\mathrm{DB\_DA}$, ζ_{DA}, and ζ_{DB} provide good performance. However, $\mathrm{DB\_DA}$ with only seven features provides the best results in terms of classification accuracies and CPU processing time.

By comparing the results reported in Tables III and V, it is easy to infer that DBFE works better than DAFE. The main reason behind this may be closely related to the fact that DAFE is not full rank (its rank is at most equal to $L-1$ where L is the number of classes). Sometimes, the aforementioned number of features is not enough in order to discriminate between different classes of interest. However, DAFE is faster than DBFE. This fact can also be observed in Tables IV and VI.

Based on our experimental results, the proposed framework improves all methods in terms of classification accuracies for Pavia University data set. For example, the proposed method improves the classification accuracy of the classification technique proposed in [39] by almost 11%. Based on the results reported in [41], the proposed method improves the OA of the previous method with PCA by almost 21% and ICA by 3.5%. These are quite important achievements from the viewpoint of classification accuracy (in this regard, our framework provides some of the best classification results ever reported in the literature for the considered scene). The main disadvantage of the proposed method is the fact that the final result is dependent on the second FE and it is difficult to anticipate which one of ζ_{DA} or ζ_{DB} works better. The investigation of these aspects will be a subject for our future research efforts.

TABLE V
PAVIA UNIVERSITY: CLASSIFICATION ACCURACIES OBTAINED BY DIFFERENT METHODS FOR THE ROSIS PAVIA UNIVERSITY SCENE AFTER APPLYING DBFE

Class Number	Class Name	Spec (29)	AP (1479)	DB_DA (7)	DB_DB (30)	ζ_{DA} (36)	ζ_{DB} (59)
1	Asphalt	85.0	**98.1**	97.8	97.3	97.6	96.7
2	Meadows	68.0	94.4	**97.4**	88.8	96.9	95.8
3	Gravel	69.4	98.0	97.2	74.7	**98.3**	87.0
4	Trees	95.2	87.3	98.3	95.2	98.5	**99.3**
5	Metal sheets	99.8	99.6	99.5	**99.9**	**99.9**	99.8
6	Soil	93.6	**100**	99.9	99.4	**100**	99.9
7	Bitumen	86.1	**100**	99.6	99.7	99.5	99.9
8	Bricks	87.3	98.1	99.2	98.9	99.3	**99.4**
9	Shadows	**97.5**	97.1	94.0	95.9	95.7	91.8
Kappa	–	0.7441	0.9481	**0.9746**	0.9078	0.9732	0.9576
OA	–	79.56	96.04	**98.07**	92.91	97.97	96.78
AA	–	86.91	96.98	98.15	94.47	**98.46**	96.67

The number of features used for classification purposes is reported in the parentheses.

TABLE VI
PAVIA UNIVERSITY: CPU PROCESSING TIME (IN SECONDS) OF DIFFERENT METHODS AFTER APPLYING DBFE

Spec	AP	DB_DA	DB_DB	ζ_{DA}	ζ_{DB}
36	478	201	803	833	827

TABLE VII
INDIAN PINES: CLASSIFICATION ACCURACIES OBTAINED BY DIFFERENT METHODS FOR THE AVIRIS INDIAN PINES SCENE AFTER APPLYING DAFE

Class Number	Class Name	Spec (13)	AP (663)	DA_DA (13)	DA_DB (45)	ζ_{DA} (26)	ζ_{DB} (58)
1	Corn-notil	54.3	82.7	82.8	76.1	**88.5**	76.2
2	Corn-mintill	52.8	**96.0**	95.5	90.0	95.1	88.2
3	Corn	67.9	92.9	**99.4**	96.7	98.9	95.1
4	Grass-pasture	89.9	93.7	**95.7**	94.4	94.6	94.6
5	Grass-trees	87.9	96.1	**97.5**	95.9	97.1	95.4
6	Hay-windrowed	97.0	**99.7**	98.6	99.3	98.6	99.3
7	Soybean-notill	63.8	**91.6**	81.0	87.0	86.6	83.5
8	Soybean-mintill	44.8	85.1	76.7	73.4	**91.3**	73.0
9	Soybean-clean	64.3	87.7	**89.8**	89.3	89.7	86.7
10	Wheat	98.1	99.3	99.3	**100**	99.3	**100**
11	Woods	85.2	99.3	**99.9**	91.8	99.4	93.8
12	Bldg-grass-tree-drives	80.9	99.0	**99.3**	98.7	**99.3**	97.2
13	Stone-steel-towers	93.3	**100**	**100**	**100**	**100**	**100**
14	Alfalfa	56.4	**97.4**	94.8	**97.4**	94.8	**97.4**
15	Grass-pasture-mowed	**100**	**100**	**100**	90.9	**100**	**100**
16	Oats	80.0	**100**	**100**	**100**	**100**	**100**
Kappa	–	0.6118	0.8987	0.8683	0.8359	**0.9227**	0.8295
OA	–	65.47	91.13	88.47	85.53	**93.27**	84.99
AA	–	76.07	95.07	94.44	92.59	**95.86**	92.56

The number of features used for classification purposes is reported in the parentheses.

TABLE VIII
INDIAN PINES: CPU PROCESSING TIME (IN SECONDS) OF DIFFERENT METHODS AFTER APPLYING DAFE

Spec	AP	DA_DA	DA_DB	ζ_{DA}	ζ_{DB}
2	17	13	64	13	65

Fig. 5 shows classification maps for different methods started by DAFE applied on Pavia University.

2) Indian Pines: The low spatial resolution of this data set adds more complexity, since it leads to the presence of highly mixed pixels (which are mainly due to the early growth cycle of most of the agricultural features in the scene). In this case, the classification results may be degraded by the presence of mixed pixels in the scene. In addition, the significant

TABLE IX
INDIAN PINES: CLASSIFICATION ACCURACIES OBTAINED BY DIFFERENT METHODS FOR THE AVIRIS INDIAN PINES SCENE AFTER APPLYING DBFE

Class Number	Class Name	Spec (16)	AP (816)	DB_DA (13)	DB_DB (43)	ζ_{DA} (29)	ζ_{DB} (59)
1	Corn-notil	51.4	79.9	71.6	75.2	**86.2**	73.7
2	Corn-mintill	56.8	96.4	96.5	90.0	**96.6**	90.4
3	Corn	72.8	88.5	**94.0**	**94.0**	**94.0**	**94.0**
4	Grass-pasture	85.2	93.5	**95.5**	93.7	95.0	93.2
5	Grass-trees	87.9	**99.0**	95.9	98.4	94.8	98.5
6	Hay-windrowed	94.3	**99.3**	99.0	99.0	99.0	99.0
7	Soybean-notill	63.6	**87.2**	78.1	80.0	86.8	77.8
8	Soybean-mintill	46.9	82.0	75.2	73.8	**85.9**	70.3
9	Soybean-clean	65.4	84.9	**87.2**	78.1	84.9	77.6
10	Wheat	99.3	**100**	98.7	**100**	**100**	**100**
11	Woods	81.1	**99.6**	99.5	94.6	99.4	93.1
12	Bldg-grass-tree-drives	75.4	98.7	**99.7**	99.3	**99.7**	98.4
13	Stone-steel-towers	91.1	**100**	**100**	**100**	**100**	**100**
14	Alfalfa	69.2	**97.4**	**97.4**	**97.4**	**97.4**	**97.4**
15	Grass-pasture-mowed	81.8	**100**	**100**	**100**	**100**	**100**
16	Oats	40.0	**100**	**100**	80.0	**100**	80.0
Kappa	–	0.6058	0.8808	0.8388	0.8254	**0.9001**	0.8083
OA	–	64.99	89.56	85.90	84.64	**91.27**	83.11
AA	–	72.67	94.18	93.05	90.88	**95.01**	90.25

The number of features used for classification purposes is reported in the parentheses.

TABLE X
INDIAN PINES: CPU PROCESSING TIME (IN SECONDS) OF DIFFERENT METHODS AFTER APPLYING DBFE

Spec	AP	DB_DA	DB_DB	ζ_{DA}	ζ_{DB}
26	45	44	132	45	133

differences in the number of pixels in the reference data for different classes make the classification task even more complicated.

In these data, there is a high confusion between classes Soybean-mintill and corn-notill which degrades the class accuracies of both of them. By comparing Tables II and VII, it is easy to infer that, by performing DAFE on the input data and choosing the first features with cumulative eigenvalues above 99%, OA is reduced from 70.24% (Raw) to 65.47% (Spectral). This reveals that of only 13 features are not sufficient to discriminate between different classes, as compared to hundreds of spectral bands from the input data.

As it can be seen from Table VII, AP improves the overall accuracy of Raw in more than 25%. The main reason behind this significant improvement is that AP not only considers the spectral information, but also can model the spatial information contained in the input data. The best classification accuracies in Table VII are achieved by the proposed method; ζ_{DA}, which improves the overall accuracy of Spectral, AP, DA_DA, DA_DB, and ζ_{DB} by almost 28, 2, 5%, 8%, and 9%, respectively. It should be noted that the new method can discriminate different classes by considering only 26 features. Moreover, the CPU processing time of the proposed method is acceptable and takes only 13 s to classify the input data set in the considered computing environment. Table VIII gives information regarding the CPU processing time of different methods after applying DAFE.

After ζ_{DA}, AP exhibits the best performance among other techniques in terms of classification accuracies. This confirms that the consideration of spatial information has a significant influence on the discrimination of different classes. By including a second FE step, although classification accuracy for some classes such as classes 3 and 4 are improved, the overall accuracy of AP is reduced from 91.13% to 88.47% (DA_DA) and 85.53% (DA_DB).

Table IX gives information related to the classification accuracies of different methods after DBFE. The corresponding CPU processing times are listed in Table X. By comparing Tables II and IX, one can infer that the OA of the Raw classification decreases when DBFE is performed. Again, the proposed method outperforms other techniques with acceptable CPU processing time (45 s) in this particular case.

In contrast, it is also important to emphasize AP exhibits an acceptable performance in terms of classification accuracies when compared to other classifiers (its performance is only slightly lower than ζ_{DA}). AP provides 720 features. This reveals that RF is a robust classifier when dealing with very high-dimensional data. Also, it is worth mentioning that ζ_{DA} provides the best performance overall, and improves the OA of Spectral, AP, DB_DA, DB_DB, and ζ_{DB} by more than 26%, 1.7%, 6.2%, 6.5%, and 7.9%, respectively.

As can be seen from Tables VII and IX, AP provides 585 and 720 features, respectively. The table shows that RF can properly handle classification problems consisting of high-dimensional input features and limited training samples, with acceptable CPU processing time. In almost all cases, DAFE outperforms DBFE in terms of classification accuracies and CPU processing time. A possible reason for this may be

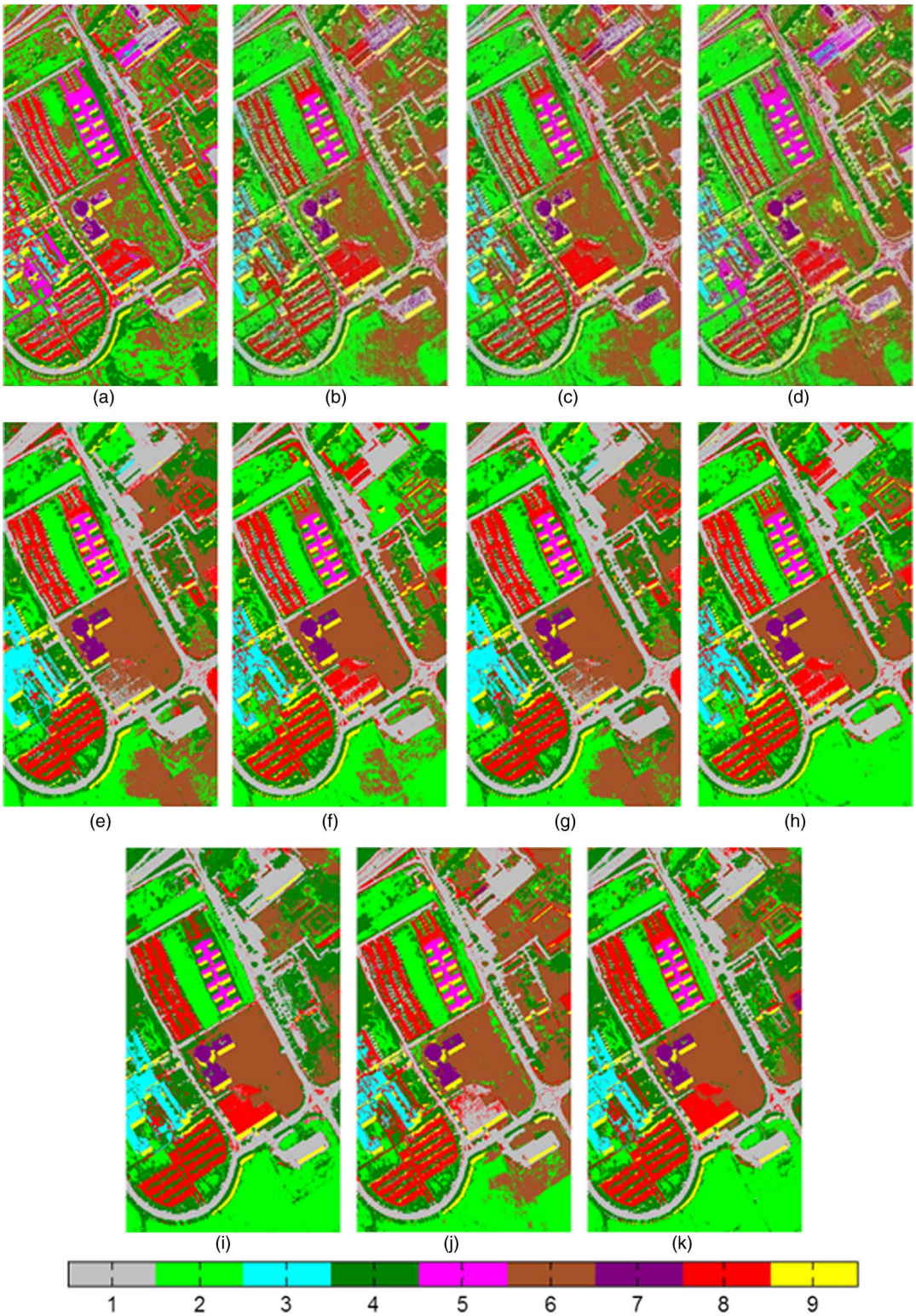

Fig. 5. Pavia University: (a)–(f), classification maps for different methods started by DAFE: (a) Raw, (b) AP, (c) DA_DA, (d) DA_DB, (e) ζ_{DA}, and (f) ζ_{DB}. (g)–(k), classification maps of different methods started by DBFE: (g) AP, (h) DB_DA, (i) DB_DB, (j) ζ_{DBA}, and (k) ζ_{DB}.

the fact that the number of selected features used by DBFE is not sufficient. As a result, more features need to be considered in order to provide more consistent results in the case of DBFE, which can be computationally intensive and its performance is highly dependent on the training samples.

Fig. 6 shows classification maps for different methods started by DAFE applied on Indian Pines.

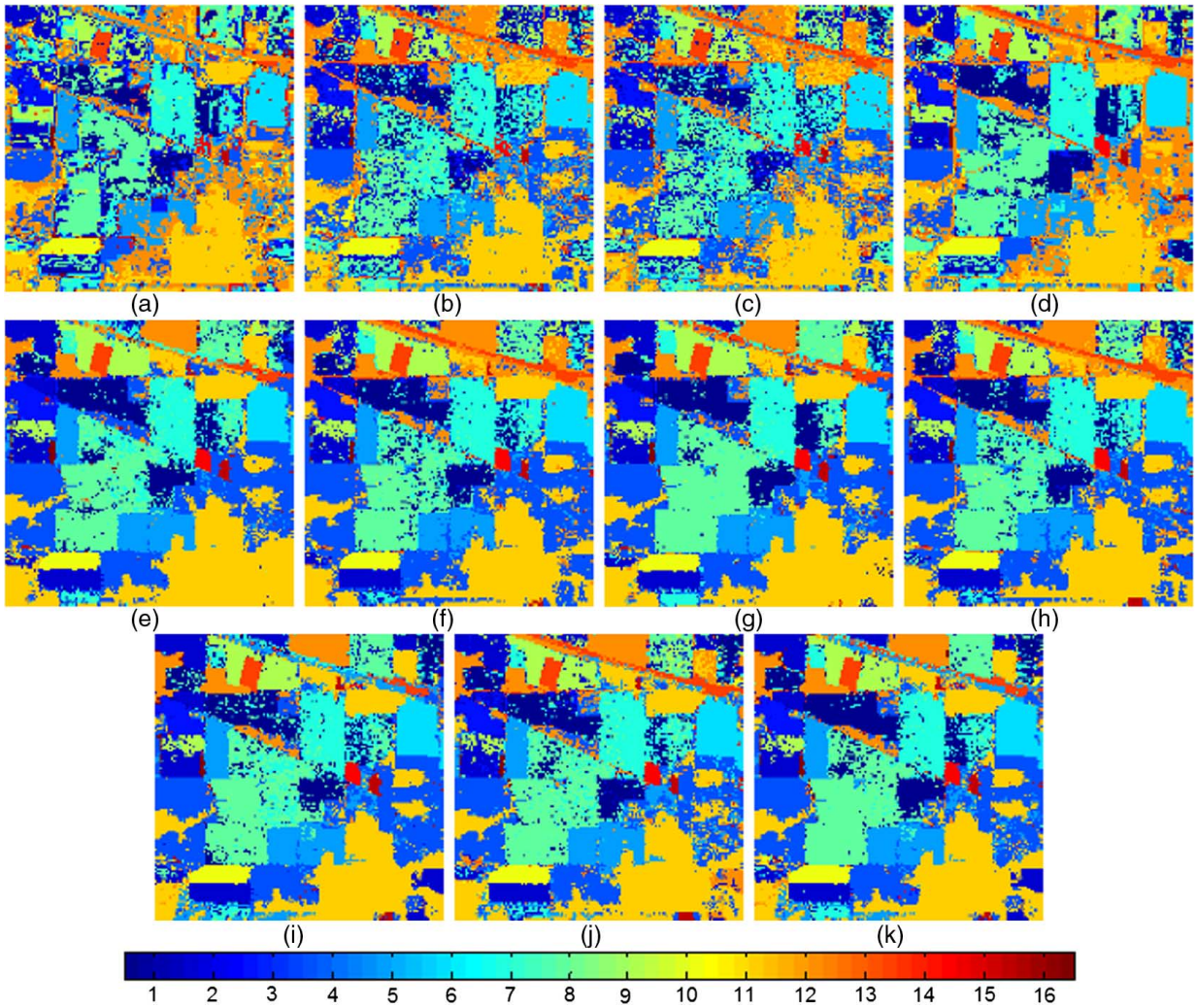

Fig. 6. AVIRIS (a)–(f), classification maps for different methods after applying DAFE: (a) Raw, (b) AP, (c) DA_DA, (d) DA_DB, (e) ζ_{DA}, and (f) ζ_{DB}. (g)–(k), classification maps of different methods after applying DBFE: (g) AP, (h) DB_DA, (i) DB_DB, (j) ζ_{DBA}, and (k) ζ_{DB}.

IV. Conclusion

In this paper, we have developed a new automatic framework for the classification of hyperspectral images. Our framework uses both spectral and spatial information. In order to include the spatial information, morphological APs are taken into account. For reducing the redundancy of the extracted features and deal with the curse of dimensionality introduced by the Hughes effect, supervised FE methods (DAFE and DBFE) are considered. The proposed framework is extensively tested on two widely used hyperspectral data sets, i.e., the ROSIS-03 Pavia University scene and the AVIRIS Indian Pines. Different methods have been used to implement the presented framework, and the results provided have been compared in terms of classification accuracies and CPU processing time.

It should be noted that the two selected hyperspectral data sets represent very different case studies collected by different instruments. The former is related to urban area problems and presents high spatial resolution. In turn, the latter has medium-size spatial resolution and is related to agricultural land-cover classification problems. The good classification accuracies obtained in both case studies indicate the good generalization properties of the presented framework. In addition, the new approach achieves better classification accuracies than other widely used classification techniques, with acceptable CPU processing time. We emphasize that the proposed procedure is fully automatic, which is a highly desirable feature.

A topic of future investigation is the optimal selection (in terms of classification accuracies) of the FE method in the second stage of the proposed approach. Another topic deserving future research is the development of parallel implementations of the presented approach in high-performance computing architectures, although the processing times reported in our experiments (measured in a standard desktop CPU) are quite fast for the considered data sets.

Acknowledgment

The ROSIS data and corresponding reference information were kindly provided by Prof. P. Gamba from the University of Pavia, Italy.

References

[1] 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. 1, pp. S110–S122, 2009.

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

[3] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*. Hoboken, NJ, USA: Wiley, 2003.

[4] P. Ghamisi, M. S. Couceiro, and J. A. Benediktsson. "Classification of hyperspectral images with binary fractional order Darwinian PSO and random forests," in *Proc. SPIE 8892, Image Signal Process. Remote Sens. XIX,* Oct. 17, 2013, p. 88920S, doi: 10.1117/12.2027641.

[5] S. Tadjudin, "Classification of high dimensional data with limited training samples," Ph.D. thesis, School of Electrical and Computer Engineering, Purdue Univ., West Lafayette, IN, 1998, 123pp. [Online]. Available: http://dynamo.ecn.purdue.edu/~landgreb/publications.html.

[6] 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.

[7] H. Derin and P. A. Kelly, "Discrete-index Markov-type random processes," *Proc. IEEE*, vol. 77, no. 10, pp. 1485–1510, Oct. 1989.

[8] F. Bovolo and L. Bruzzone, "A context-sensitive technique based on support vector machines for image classification," in *Proc. Pattern Recogn. Mach. Intell.,* 2005, pp. 260–265.

[9] P. Ghamisi, J. A. Benediktsson, and M. O. Ulfarsson, "Spectral–spatial classification of hyperspectral images based on hidden Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. PP, no. 99, pp. 1–10, Jun. 2013 [Online]. Available: http://dx.doi.org/10.1109/TGRS.2013.2263282.

[10] P. Ghamisi, M. S. Couceiro, N. M. F. Ferreira, and L. Kumar, "Use of Darwinian particle swarm optimization technique for the segmentation of remote sensing images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.,* 2012, pp. 4295–4298.

[11] P. Ghamisi, M. S. Couceiro, J. A. Benediktsson, and N. M. F. Ferreira, "An efficient method for segmentation of images based on fractional calculus and natural selection," *Expert Syst. Appl.*, vol. 39, no. 16, pp. 12 407–12 417, 2012.

[12] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral–spatial classification of hyperspectral imagery based on partitional clustering techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2973–2987, Aug. 2009.

[13] P. Ghamisi, M. S. Couceiro, F. M. L. Martins, and J. A. Benediktsson, Multilevel Image Segmentation Based on Fractional-Order Darwinian Particle Swarm Optimization, *IEEE Trans. Geosci. Remote Sens.*, vol. PP, no. 99, pp. 1–13, Jun. 2013, doi: 10.1109/TGRS.2013.2260552 [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6524014&isnumber=4358825.

[14] P. Ghamisi, M. S. Couceiro, M. Fauvel, and J. A. Benediktsson, "Integration of segmentation techniques for classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 342–346, Jan. 2014.

[15] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 309–320, Feb. 2001.

[16] J. A. Palmason, J. A. Benediktsson, J. R. Sveinsson, and J. Chanussot, "Classification of hyperspectral data from urban areas using morphological preprocessing and independent component analysis," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, no. 3, 2005, pp. 176–179.

[17] M. Dalla Mura, J. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, Oct. 2010.

[18] M. Pedergnana, P. Marpu, M. Dalla Mura, J. Benediktsson, and L. Bruzzone, "A novel technique for optimal feature selection in attribute profiles based on genetic algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 6, pp. 3514–3528, Jun. 2013.

[19] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, p. 532, 2001.

[20] L. Breiman, "RF tools a class of two eyed algorithms," in *Proc. SIAM Workshop*, San Francisco, CA, USA, 2003.

[21] P. Marpu, M. Pedergnana, M. Dalla Mura, J. A. Benediktsson, and L. Bruzzone, "Automatic generation of standard deviation attribute profiles for spectral–spatial classification of remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 2, pp. 293–297, Mar. 2013.

[22] C. Lee and D. A. Landgrebe, "Decision boundary feature extraction for non-parametric classification," *IEEE Trans. Syst. Man Cybern.*, vol. 23, no. 2, pp. 433–444, Mar./Apr. 1993.

[23] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. New York, NY, USA: Academic Press, 1974.

[24] L. Jimenez and D. A. Landgrebe, "Hyperspectral data analysis and supervised feature reduction via projection pursuit," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 6, pp. 2653–2667, Nov. 1999.

[25] C. Lee and D. A. Landgrebe, "Feature extraction based on decision boundaries," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 4, pp. 388–400, Apr. 1993.

[26] J. Serra, *Mathematical Morphology, Theoretical Advances*, vol. 2, New York, NY, USA: Academic Press, 1988.

[27] J. Serra, *Image Analysis and Mathematical Morphology.* London, U.K.: Academic Press, 1982.

[28] P. Soille, "Morphological Image Analysis, Principles and Applications," vol. 2, Berlin, Germany: Springer Verlag, 2003.

[29] L. Najman and H. Talbot, *Mathematical Morphology*, Hoboken, NJ, USA: Wiley, 2010.

[30] E. J. Breen and R. Jones, "Attribute openings, thinnings, and granulometries," *Comput. Vision Image Understand.*, vol. 64, no. 3, pp. 377–389, 1996.

[31] M. D. Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute filters for the analysis of very high resolution remote sensing images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.* (IGARSS 2009), vol. 3, 2009, pp. III-97–III-100.

[32] P. Salembier, A. Oliveras, and L. Garrido, "Antiextensive connected operators for image and sequence processing," *IEEE Trans. Image Process.*, vol. 7, no. 4, pp. 555–570, Apr. 1998.

[33] V. Caselles and P. Monasse, *Geometric Description of Images as Topographic Maps*, 1st ed. New York, NY, USA: Springer Publishing Company, Incorporated, 2009.

[34] E. R. Urbach, J. B.T.M. Roerdink, and M. H. F. Wilkinson, "Connected shape-size pattern spectra for rotation and scale-invariant classification of gray-scale images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 2, pp. 272–285, Feb. 2007.

[35] X. Huang and L. Zhang, "An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 257–272, Jan. 2013.

[36] M. Dalla Mura, "Advanced techniques based on mathematical morphology for the analysis of remote sensing images," Ph.D. dissertation, Univ. Trento, Trento, Italy, 2011.

[37] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis*, 4th ed. Berlin, Germany: Springer Heidelberg, 2006.

[38] P. Ghamisi, J. A. Benediktsson, and J. R. Sveinsson, "Automatic spectral–spatial classification framework based on attribute profiles and supervised feature extraction," *IEEE Trans. Remote Sens. Geosci.*, vol. PP, no. 99, pp. 1–12, Dec. 2013.

[39] 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, Nov. 2008.

[40] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson, "SVM- and MRF-based method for accurate classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 736–740, Oct. 2010.

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

Pedram Ghamisi (S'13) graduated with the B.Sc. degree in civil (survey) engineering from the Tehran South Campus, Azad University, Tehran, Iran. Then, he recieved the M.Sc. degree in remote sensing from K.N. Toosi University of Technology, Tehran, Iran, in 2012. He is currently a Ph.D. student in electrical and computer engineering at the University of Iceland, Reykjavík, Iceland.

His research interests are in remote sensing and image analysis with the current focus on spectral and spatial techniques for hyperspectral image classification.

Mr. Ghamisi received the Best Researcher Award for M.Sc. students in K. N. Toosi University of Technology in the academic year 2010–2011. He was the recipient of the IEEE Mikio Takagi Prize, which was awarded for the first place in the Student Paper Competition at the 2013 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Melbourne, Australia, July 2013. He serves as a Reviewer for a number of journals including IEEE TGRS, IEEE Trans. Image Processing, IEEE JSTARS and IEEE GRSL.

Jón Atli Benediktsson (S'84–M'90–SM'99–F'04) received the Cand.Sci. degree in electrical engineering from the University of Iceland, Reykjavik, Iceland, in 1984, and the M.S.E.E. and Ph.D. degrees from Purdue University, West Lafayette, IN, USA, in 1987 and 1990, respectively.

Currently, he is a Pro Rector for Academic Affairs and Professor of electrical and computer engineering with the University of Iceland. His research interests are in remote sensing, biomedical analysis of signals, pattern recognition, image processing, and signal processing, and he has published extensively in those fields.

From 2011 to 2012, he was a President of the IEEE Geoscience and Remote Sensing Society (GRSS) and has been on the GRSS AdCom since 2000. From 2003 to 2008, he was an Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS) and has served as an Associate Editor of TGRS since 1999, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS since 2003, and IEEE Access since 2013. From 2007 to 2010, he is in the International Editorial Board of the *International Journal of Image and Data Fusion* and was the Chairman of the Steering Committee of IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (J-STARS). He is a Co-founder of the biomedical start up company Oxymap (www.oxymap.com). He is a Fellow of SPIE.

Dr. Benediktsson received the Stevan J. Kristof Award from Purdue University, in 1991 for outstanding graduate student in remote sensing. In 1997, he was the recipient of the Icelandic Research Council's Outstanding Young Researcher Award; in 2000, he was granted the IEEE Third Millennium Medal; in 2004, he was a co-recipient of the University of Iceland's Technology Innovation Award; in 2006, he received the yearly research award from the Engineering Research Institute of the University of Iceland; and in 2007, he received the Outstanding Service Award from the IEEE GEOSCIENCE AND REMOTE SENSING SOCIETY. He is co-recipient of the 2012 IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING Paper Award. He received the 2013 IEEE/VFI Electrical Engineer of the Year Award and in 2013, he was a co-recipient of the IEEE GRSS Highest Impact Paper Award. He is a Member of the Association of Chartered Engineers in Iceland (VFI), Societas Scinetiarum Islandica and Tau Beta Pi.

Gabriele Cavallaro received the B.S. and M.S. degrees in telecommunications engineering from the University of Trento, Trento, Italy, in 2011 and 2013, respectively. He did the Master thesis at the University of Iceland, Reykjavik, Iceland, in remote sensing field on morphological attribute filters based on the Inclusion Tree for the analysis of very high-resolution remote sensing images. At present, he is a Ph.D. student at the University of Iceland and at the University of Extremadura, Cáceres, Spain.

Antonio Plaza (M'05–SM'07) received the M.S. and Ph.D. degrees in computer engineering from the University of Extremadura, Cáceres, Spain.

He is a Associate Professor (with accreditation for Full Professor) with the Department of Technology of Computers and Communications, University of Extremadura, where he is the Head of the Hyperspectral Computing Laboratory (HyperComp). From 2007 to 2011, he was the Coordinator of the Hyperspectral Imaging Network, a European project with total funding of 2.8 MEuro. He authored more than 370 publications, including more than 100 JCR journal papers (60 in IEEE journals), 20 book chapters, and over 230 peer-reviewed conference proceeding papers (90 in IEEE conferences). He has guest edited seven special issues on JCR journals (three in IEEE journals). He has been a Chair for the IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (2011).

Dr. Plaza is a recipient of the recognition of Best Reviewers of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, in 2009, and a recipient of the recognition of Best Reviewers of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, in 2010 and a journal for which he has served as Associate Editor in 2007–2012. From 2011 to 2012, he was a Member of the Editorial Board of the IEEE GEOSCIENCE AND REMOTE SENSING NEWSLETTER and a member of the steering committee of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING in 2012. He is also an Associate Editor for the IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE. From 2011 to 2012, he served as the Director of Education Activities for the IEEE GEOSCIENCE AND REMOTE SENSING SOCIETY (GRSS) and is currently serving as a President of the Spanish Chapter of IEEE GRSS (since November 2012). He is currently serving as the Editor-in-Chief of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING journal, since January 2013. Additional information: http://www.umbc.edu/rssipl/people/aplaza.