Complementarity of Discriminative Classifiers and Spectral Unmixing Techniques for the Interpretation of Hyperspectral Images

Jun Li, *Member, IEEE*, Inmaculada Dópido, Paolo Gamba, *Fellow, IEEE*, and Antonio Plaza, *Senior Member, IEEE*

Abstract—Classification and spectral unmixing are two important techniques for hyperspectral data exploitation. Traditionally, these techniques have been exploited independently. In this paper, we propose a new technique that exploits their complementarity. Specifically, we develop a new framework for semisupervised hyperspectral image classification that naturally integrates the information provided by discriminative classification and spectral unmixing. The idea is to assign more confidence to the information provided by discriminative classification for those pixels that can be easily catalogued due to their spectral purity. For those pixels that are more highly mixed in nature, we assign more confidence to the information provided by spectral unmixing. In this case, we use a traditional spectral unmixing chain to produce the abundance fractions of the pure signatures (endmembers) that model the mixture information at a subpixel level. The decision on which source of information is prioritized in the process is taken adaptively, when new unlabeled samples are selected and included in our semisupervised framework. In this regard, the proposed approach can adaptively integrate these two sources of information without the need to establish any weight parameters, thus exploiting the complementarity of classification and unmixing and selecting the most appropriate source of information in each case. In order to test our concept, which has similar computational complexity as traditional semisupervised classification strategies, we have used two different hyperspectral data sets with different characteristics and spatial resolution. In our experiments, we consider two different discriminative classifiers: multinomial logistic regression and probabilistic support vector machine. The obtained results indicate that the proposed approach, which jointly exploits the features provided by classification and spectral unmixing in adaptive fashion, offers an effective solution to improve classification performance in hyperspectral scenes containing mixed pixels.

Index Terms—Discriminative classification, hyperspectral imaging, semisupervised learning, spectral unmixing.

I. INTRODUCTION

HYPERSPECTRAL imaging is a widely used technique in remote sensing that collects hundreds of images, at different wavelength channels, for the same area on the surface of the Earth [1]. The special characteristics of hyperspectral data sets pose different processing problems [2], [3], which must be necessarily tackled under specific mathematical formalisms, such as classification [4] and spectral unmixing [5], among others [6].

Hyperspectral image classification has been a very active area of research in the last few years [7], [8]. Given a set of observations (i.e., possibly mixed pixel vectors), the goal of classification is to assign a unique label to each pixel vector so that it is well defined by a given class [9]. Discriminative approaches have been widely used in hyperspectral image classification, as they are usually less complex and can generally obtain better performance than generative models in the presence of limited training samples [9]–[12]. For instance, the support vector machine (SVM) [13], [14] and the multinomial logistic regression (MLR) [15], [16] are among the most widely used discriminative classifiers in hyperspectral analysis, since these methods can produce sparse solutions and work effectively in a nonlinear (or kernel) space to improve class separability [17].

Under the discriminative model, both supervised and semisupervised approaches for hyperspectral image classification have been developed. As shown by previous studies, supervised classification is generally a difficult task due to the unbalance between the high dimensionality of the spectral vectors and the limited availability of (labeled) training samples *a priori* [1]. This has fostered the development of semisupervised techniques that are able to exploit unlabeled training samples, which can be obtained from a (limited) set of available labeled samples without significant effort/cost [18]. For instance, in [19], transductive SVMs are used to gradually search a reliable separating hyperplane (in the kernel space) with a transductive process that incorporates both labeled and unlabeled samples in the training phase. In [20], a semisupervised method is presented that exploits the wealth of unlabeled samples in the image and naturally gives relative importance to the labeled ones through a graph-based methodology. In [21], a semisupervised SVM is presented that exploits the wealth of unlabeled samples for regularizing the training kernel representation locally by means of cluster kernels. A related idea is explored in [22] in the context of urban classification. In [23], a new semisupervised approach is presented that exploits unlabeled training samples (selected by means of an active selection strategy based on the entropy of

the samples). Here, unlabeled samples are used to improve the estimation of the class distributions. In [24], a novel context-sensitive semisupervised SVM is presented that exploits the contextual information of the pixels belonging to the neighborhood system of each training sample in the learning phase to improve the robustness to possible mislabeled training patterns. In [25], a semisupervised self-learning technique (which can exploit either the SVM or the MLR classifier) is shown to be able to select the most useful and informative unlabeled samples for classification purposes. This method is based on the adaptation of active learning methods (in which a trained expert actively selects unlabeled samples [26]).

An important problem for hyperspectral image classification is the presence of mixed pixels [27]. If a pixel is highly mixed, it is very difficult to determine its class label as it is not easily separable from other classes. A common way for dealing with mixed pixels has been spectral unmixing [5], which aims at estimating the abundance fractions of a set of pure spectral signatures (called *endmembers* in unmixing jargon) that can be then used to determine endmember proportions within the pixel [28]. The mixing effects arise when distinct materials are combined into a homogeneous or an intimate mixture, which occurs independently of the spatial resolution of the sensor [29]. To address these issues, several techniques for endmember identification [30], [31] and abundance estimation [32] have been developed in the literature under the assumption that each pixel vector measures the response of multiple underlying materials. Despite the very important advances reported in spectral unmixing literature, these approaches still find challenges such as the determination of the number of spectral endmembers [33] or the proper identification of the spectral signatures associated to pure spectral components when the scene is dominated by mixed pixels [34]. This has recently fostered the idea of jointly exploiting spectral unmixing and classification for advanced hyperspectral interpretation.

In order to take advantage of the complementary properties of spectral unmixing and classification, several approaches have been studied in the recent literature for the integration of these two techniques. This is important, as these techniques have traditionally evolved independently, although they exhibit many complementary properties. For instance, in [35], spectral unmixing was used as a feature extraction strategy prior to classification. It was found that spectral unmixing can be used to derive suitable features (i.e., abundance maps [36]) for classification purposes. This idea was further explored in [37], in which several spectral unmixing chains (unsupervised and supervised) were used to derive suitable features for classification purposes. A limitation found in this strategy is that the information provided by spectral unmixing is exploited prior to classification, whereas in certain scenarios, it is important to take advantage of both sources of information in complementary fashion. More recently, the synergistic nature of spectral unmixing and classification has been explored in the context of a semisupervised framework [38]. This strategy provides an approach for hyperspectral data interpretation that considers the information provided by spectral unmixing in order to select new training samples for classification purposes, where the relative importance of classification and unmixing can be

controlled by the end user. In this approach, there is a need to set a parameter that defines the relative weight of classification to unmixing (and vice versa). The setting of this parameter may not be straightforward in practice and remains equal for all the pixels in the scene. In this regard, all the aforementioned strategies [35], [37], [38] treat all the pixels in the same way, regardless of the fact that there may be pixels that are not highly mixed (in which case a discriminative classification step may be sufficient to characterize them), whereas other pixels may be highly mixed (here, spectral unmixing can be particularly useful to enhance the classification).

In this paper, we develop a new strategy for the complementary integration of discriminative classification techniques and spectral unmixing, which offers several important contributions with regard to previous developments in this area. In the following, we summarize the main contributions of this paper with regard to previous efforts in [35], [37], and [38].

- First and foremost, the proposed approach considers separately the pixels that have high confidence to be characterized by discriminative classification approaches and the pixels that are more highly mixed in nature. This is important, as in some cases the information provided by the discriminative classification is enough to fully characterize the pixels (in particular, if they are more spectrally pure). In this case, we assign more confidence to the pixels that are discriminable, whereas the pixels that are not separable are handled by relying more on the information provided by spectral unmixing. In other words, if a pixel is well discriminated at the classification stage, we rely more on the information provided by the discriminative classifier. For those samples that cannot be well characterized by a discriminative classifier, we rely more on the information provided by abundance fractions in order to perform the classification.

- Another important contribution of the proposed approach is that it does not include any weight parameter to determine the importance of classification and unmixing. In previous works, such as [38], a weight had to be established in advance. Such weight is the same for the scene as a whole, and the pixels were not treated adaptively in terms of their information. In our newly proposed approach, the pixels with high confidence in the discriminative classification are handled by the classification stage, whereas spectral unmixing information is used when the confidence in the classification is not high. This means that the relative importance of classification and unmixing is established adaptively for each pixel. This is accomplished due to the semisupervised nature of our method, in which the unlabeled samples are properly generated by giving more importance to classification or to unmixing as required. In fact, our proposed semisupervised approach uses both labeled and unlabeled samples to train the discriminative classifier, whereas active learning techniques are used to select the most informative unlabeled training samples automatically (in self-learning fashion) by intelligently exploiting the information provided by classification or spectral unmixing.

- Last but not the least, an important characteristic of our proposed method is that it naturally (and adaptively) integrates the information provided by classification and spectral unmixing. As opposed to the work in [35], both sources of information receive equal importance *a priori* (without any weight parameters) and are exploited at the same time as needed, i.e., spectral unmixing is used in those cases in which the discriminative classification does not offer a sufficiently confident source of information to categorize a given pixel. In this case, spectral unmixing is used as an alternative source of information. This provides a more adaptive and balanced approach with regard to the methodologies presented in [35], [37], and [38]. At the same time, the proposed approach remains within similar computational complexity as the classic semisupervised discriminative classification. This is because the spectral unmixing information is fast and easy to obtain, and it can be derived in advance so that the semisupervised process can intelligently decide to resort to this source of information during the process without adding any computational burden to the overall process.

The remainder of this paper is organized as follows. Section II presents the proposed framework for complementary integration classification and spectral unmixing. Here, we particularly describe how the proposed approach has been blended with the (supervised and semisupervised) classifiers considered in this paper. Section III presents our experimental setting, with emphasis on describing the scenes considered in experiments and the experimental results conducted with such hyperspectral scenes in order to validate the presented technique. Section IV concludes with some remarks and hints at plausible future research lines.

II. PROPOSED APPROACH

Here, we describe the proposed complementary approach for exploiting discriminative classification and spectral unmixing. Let $\mathbf{X} \equiv (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathbb{R}^{d \times n}$ denote a hyperspectral image, where n is the number of pixels, and d is the number of bands in each spectral vector. Let k and p be the number of classes and spectral endmembers contained in $\mathbf{X}$, respectively. Since our final objective is to obtain a classification map $\mathbf{y} \equiv \{y_1, \ldots, y_n\}$, we assume that the number of classes and the number of endmembers are the same, i.e., $k = p$. In the remainder of this paper, we use p to denote both the number of classes and endmembers for simplicity. Let $\mathcal{X}_c = \{\mathbf{x}_1, \ldots, \mathbf{x}_t\}$ denote a set of pixels (samples) that are spectrally pure in nature and, therefore, can be modeled using the information provided by a discriminative classifier. Similarly, let $\mathcal{X}_m = \{\mathbf{x}_{t+1}, \ldots, \mathbf{x}_n\}$ denote a set of pixels (samples) that are mixed in nature and, therefore, need to be modeled using spectral unmixing. For a given spectral vector $\mathbf{x}_i$, let $\mathbf{p}_i \equiv [p_i^{(1)}, \ldots, p_i^{(p)}]$ be the classification probabilities obtained by a discriminative classifier, and let $\Phi_i \equiv [\Phi_i^{(1)}, \ldots, \Phi_i^{(p)}]$ be the fractional abundances obtained by spectral unmixing.

With the aforementioned notations in mind, the proposed approach can be summarized by the flowchart given in Fig. 1.
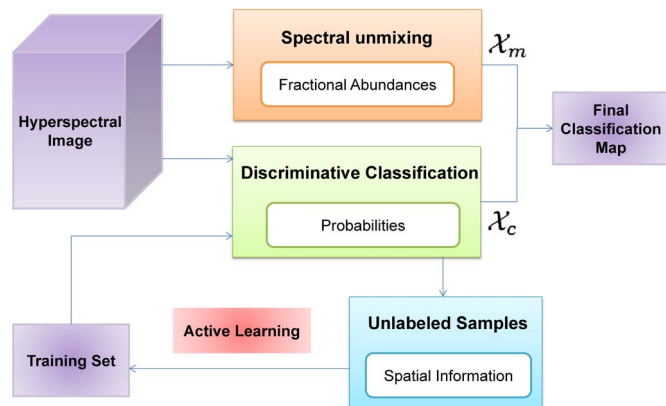

Fig. 1. Block diagram illustrating the proposed approach.

As shown in the figure, the proposed approach consists of two main steps. In the first step, we perform spectral unmixing to estimate the fractional abundances. In the second step, the classification probabilities for each pixel are obtained via discriminative classification. The final classification map is obtained by integrating the sets $\mathcal{X}_m$ and $\mathcal{X}_c$ resulting from the first and second steps, respectively. The set $\mathcal{X}_c$ can be expanded with additional unlabeled samples, which are intelligently selected using spatial information and active learning concepts. This affects the number of samples in $\mathcal{X}_m$, which are finally exploited for the generation of the final classification map. As a result, the main issue is how to identify the sets $\mathcal{X}_m$ and $\mathcal{X}_c$. In the following subsections, we provide detailed information about how to define the sets $\mathcal{X}_m$ and $\mathcal{X}_c$ and, finally, about how to exploit the complementarity of classification and spectral unmixing (in semisupervised fashion) in order to obtain the final classification map.

A. Spectral Unmixing

The strategy used in this paper for spectral unmixing is based on the well-known linear mixture model [5]. Under the linear mixture model assumption, each pixel vector $\mathbf{x}_i$ in the original scene can be modeled using the following expression:

$$\mathbf{x}_i = \sum_{e=1}^{p} \Phi_i^{(e)} \cdot \mathbf{E}^{(e)} + \mathbf{n}_i \qquad (1)$$

where $\mathbf{E}^{(e)}$ denotes the spectral response of endmember e, $\Phi_i^{(e)}$ is a scalar value designating the fractional abundance of the endmember e at the pixel $\mathbf{x}_i$, and $\mathbf{n}_i$ is a noise vector. Two physical constraints are generally imposed into the model described in (1): the abundance nonnegativity constraint (ANC), i.e., $\Phi_i^{(e)} \geq 0$, and the abundance sum-to-one constraint (ASC), i.e., $\sum_{e=1}^{p} \Phi_i^{(e)} = 1$ [39].

In this paper, we first obtain a set of endmembers from the training set using a standard endmember identification algorithm [40]. Since we are extracting the endmembers from the training set, it is possible that the number of endmembers in the whole hyperspectral image may be larger than the number of endmembers represented in the training set. At this point, it should be noted that the endmembers are obtained from the training set, and the training samples are randomly selected.

As a result, the spectral signatures obtained from the set of training samples associated to each class cannot be considered a spectrally pure signature in all cases. However, the selection of the endmembers from the training data offers important advantages in terms of computational speed (as the search for the endmembers only needs to be conducted in the labeled set) and in terms of the characterization of the endmembers (since the outcome of our proposed approach is ultimately a classification, it is more intuitive to assume that the endmembers can be used to characterize the classes). Hence, in this paper, we assumed that extracting the endmembers from the training set offers a good compromise in terms of both endmember characterization and computational performance.

Another important observation is that the ground truth that defines the labeled samples per class is generally not available for the entire hyperspectral scene. As a result, our assumption is that additional spectrally pure constituents may be available in the portions of the hyperspectral data without ground truth. In our experiments, we do not specifically ensure that each endmember is extracted from each different class in the training set, but instead let the endmember identification algorithm in [40] perform endmember extraction over the available labeled pixels in the scene. According to our experiments, in most cases, all different classes end up being represented by a different endmember in the selection process, but we do not necessarily enforce this as there may be different classes that are similar in spectral terms and could be represented by the same endmember. As the number of endmembers in the original image may be actually larger than the number of classes in the scene, in order to unmix the original image, we need to address a partial unmixing problem.

A successful technique to estimate abundance fractions in such partial unmixing scenarios is mixture-tuned matched filtering (MTMF) [41], which combines the best parts of the linear spectral unmixing model and the statistical matched filter model while avoiding some drawbacks of each parent method. From matched filtering, it inherits the ability to map a single known target without knowing the other background endmember signatures, unlike the standard linear unmixing model. From spectral mixture modeling, it inherits the leverage arising from the mixed pixel model and the constraints on feasibility, including the ASC and ANC requirements. It is essentially a target detection algorithm designed to identify the presence (or absence) of a specified material by producing a score of 1 for pixels wholly covered by the material of interest, while keeping the average score over an image as small as possible. It uses only one endmember spectrum (that of the target of interest) and therefore behaves as a partial unmixing method that suppresses background noise and estimates the subpixel abundance of a single endmember material without assuming the presence of all endmembers in the scene, as it is the case with fully constrained linear spectral unmixing [39]. If we assume that $\mathbf{E}^{(e)}$ is the endmember to be characterized, MTMF estimates the abundance fraction Φ_i of $\mathbf{E}^{(e)}$ in a specific pixel vector $\mathbf{x}_i$ of the scene as follows:

$$\hat{\Phi}_i^{(e)} = \left(\left(\mathbf{E}^{(e)T} \mathbf{R}^{-1} \mathbf{E}^{(e)} \right)^{-1} \mathbf{R}^{-1} \mathbf{E}^{(e)} \right)^T \mathbf{x}_i \qquad (2)$$

where $\mathbf{R}$ is the matrix

$$\mathbf{R} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T \qquad (3)$$

with n denoting the number of pixels in the original hyperspectral image. The outcome of this step is a set of fractional abundances $\Phi_i \equiv [\Phi_i^{(1)}, \dots, \Phi_i^{(p)}]$ for each original pixel vector $\mathbf{x}_i$.

B. Classification

In this paper, we have used two different classifiers, which provide probabilistic outputs with the form $\mathbf{p}_i \equiv [p_i^{(1)}, \dots, p_i^{(p)}]$. The first one is the MLR classifier [42] that takes advantage of a sparsity inducing prior added on the regressors in order to obtain sparse estimates [43]. In our implementation of the MLR, we use the logistic regression via splitting and augmented Lagrangian algorithm [44] to learn the MLR classifier as this technique is able to learn directly the posterior class distributions and deal with high dimensionality of hyperspectral data in a very effective way. This is particularly important when the MLR is used for semisupervised learning since, ultimately, we would like to include as many unlabeled samples as possible, a task that is difficult from the viewpoint of computational complexity. An alternative classifier considered in this paper is the probabilistic SVM in [45] and [46]. We have selected the SVM as a possible alternative to the MLR since this classifier is already widely used to analyze hyperspectral data [19], [20], whereas the MLR has only recently emerged as a feasible technique for this purpose. The LIBSVM library[1] was used in our SVM experiments.

It should be noted that both classifiers are used in this paper in both supervised and semisupervised fashion. For the semisupervised mode, we use a self-learning strategy to improve the selection of unlabeled samples [25]. Specifically, we infer the candidate set of samples $\mathcal{D}_c$ using spatial information (i.e., by applying a first-order spatial neighborhood on available labeled samples) so that high confidence can be expected in the class labels of the obtained candidate set. In a second step, we select the most informative samples from the candidate set $\mathcal{D}_c$ using an active learning algorithm (adapted here to a self-learning scenario). It should be noted that many active learning techniques are available in the literature [26], [47]. In this paper, we use the well-known breaking ties [16], [48] as a baseline to implement the proposed method. This algorithm iteratively finds new unlabeled samples minimizing the distance between the first two most probable classes.

An important issue worth being discussed at this point is the number of unlabeled samples per iteration that should be selected in order to keep a good balance between the classification accuracy and the computational cost of the algorithm. As discussed in [16] and [48], for supervised active learning, the ideal case is to include only one unlabeled sample per iteration. However, this may increase the computational cost. Another relevant issue is the spatial resolution of the data to be processed. On the one hand, when the spatial resolution is low, pixels are likely to

---

[1]http://www.csie.ntu.edu.tw/cjlin/libsvm/

be mixed, and the confidence of the candidate set is expected to be low. Therefore, one single unlabeled sample might lead to significant errors in the classification. On the contrary, a few unlabeled samples could balance the propagation error. On the other hand, when the spatial resolution is high, the classes may have better separability, and a smaller number of unlabeled samples per iteration could lead to good classification performance. Given the different possibilities available, which are also linked to the considered case study and application domain, in this paper, we have decided to test several strategies, i.e., 1 unlabeled sample per iteration, 10 unlabeled samples per iteration, and 30 unlabeled samples per iteration, to analyze the impact of the number of unlabeled samples in the final classification results.

As for the stopping criterion of the proposed approach, as discussed in [16] and [48], the ideal case is to stop the process when convergence is achieved in the active learning process. However, as shown in [16], the convergence may come with a large number of unlabeled samples. In this case, computational issues are involved. In this paper, we decided to use a fixed total number of unlabeled samples in order to simplify the execution and experimental evaluation of the proposed approach while keeping the computational complexity within manageable limits.

C. Complementary Integration

As mentioned in the previous subsection, our methodology is designed to exploit the information provided by discriminative classification and spectral unmixing in complementary fashion, by intelligently selecting the source of information that is more appropriate for each pixel. In other words, some pixels will be assigned to the set $\mathcal{X}_c$, and in this case, the information obtained by discriminative classification, for example, $\mathbf{p}_i$ for $i = 1, \ldots, t$, will be used. On the other hand, the rest of the pixels will be assigned to the set $\mathcal{X}_m$, and in this case, the information obtained by spectral unmixing, for example, Φ_j for $j = t+1, \ldots, n$, will be used. Since we are using probabilistic classifiers, their output provides a degree of confidence in the assignment of a given pixel to each of the available classes, which can be considered similar to the information provided by spectral unmixing, in which the abundances indicate the fractional coverage of each class within the given pixel. An important consideration at this point is that, if we can determine the set of $\mathcal{X}_c$, we can automatically determine the set $\mathcal{X}_m$ as they are mutually exclusive. Here, our assumption is that the pixels for which we have high confidence in the discriminative classifier should be assigned to $\mathcal{X}_c$, whereas the remaining pixels (for which we do not have enough confidence) should be assigned to $\mathcal{X}_m$.

In the previous subsection, we mentioned that the candidate set $\mathcal{D}_c$ (which provides the pool of candidates for the selection of unlabeled samples by means of active learning) is expected to contain samples that are labeled with high confidence. This is because these samples fulfill both a global and a local criterion, as the samples are in the spatial proximity of other samples that have been labeled with high confidence by the probabilistic classifier. As a result, the samples in $\mathcal{D}_c$ are expected to be properly modeled by the information provided by the discrimi-
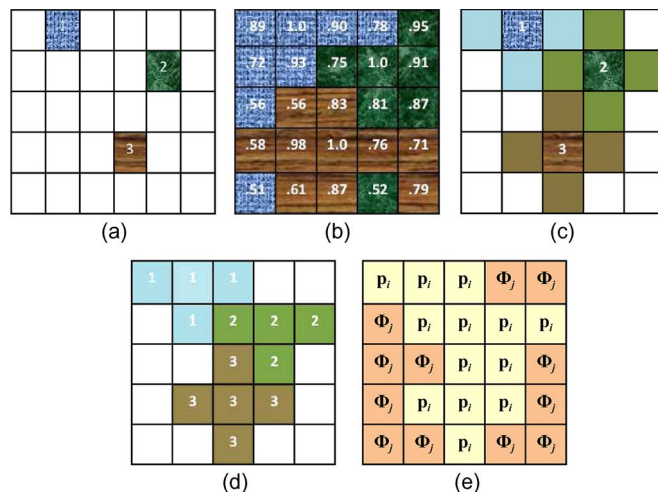

Fig. 2. Toy example illustrating the procedure used to select the source of information (classification or unmixing) for each pixel. (a) Labeled samples. (b) Probabilistic classification. (c) First-order neighborhood. (d) Set of candidates $\mathcal{D}_c$. (e) Source of information for each pixel.

native classification. In turn, the remaining samples are less confident, and we cannot guarantee that they can be effectively characterized by the discriminative classifier. In this case, we use spectral unmixing as the main source of information to characterize these pixels.

For illustrative purposes, Fig. 2 shows a toy example illustrating the procedure used to select the most appropriate source of information for each pixel. In Fig. 2(a), we display the original image and the available labeled samples for three different classes, labeled as 1, 2, and 3. These samples are used to train a probabilistic classifier, which provides a probabilistic output for each pixel, as illustrated in Fig. 2(b). In Fig. 2(c), we use a first-order neighborhood system to determine the samples in the proximity of the most confident samples in Fig. 2(b). Fig. 2(d) shows the candidate set $\mathcal{D}_c$. It should be noted that not all the samples in Fig. 2(c) are included in Fig. 2(d) as the probabilistic information in Fig. 2(b) can be used to remove some of the neighboring samples, according to the aforementioned global and local criterion. Finally, Fig. 2(e) shows the source of information (classification or spectral unmixing) used for each pixel. In the case of samples included in $\mathcal{D}_c$, we use the probabilities, i.e., $\mathbf{p}_i$, as these samples exhibit high confidence in the classification on both a local and a global level (the local level comes from the fact that these pixels are in the first-order spatial neighborhood, whereas the global level comes from the fact that the pixels have been assigned to a class by the global discriminative classifier). For the samples that are not included in $\mathcal{D}_c$, we use the abundance fractions Φ_j, as we cannot guarantee that discriminative classification will be able to properly characterize these samples. Based on this observation, during the active learning iterations, we assign the samples in the candidate pool $\mathcal{D}_c$ to the discriminative set $\mathcal{X}_c$ and the remaining samples to the spectral unmixing set $\mathcal{X}_m$. Furthermore, it is interesting to observe that the information from two sources is integrated in the form of chessboard. This is expected as we iteratively enlarge the candidate pool $\mathcal{D}_c$. At this point, it should be noted that, as the choice between classification and unmixing is adaptive and the unmixing part has no straightforward link to the active
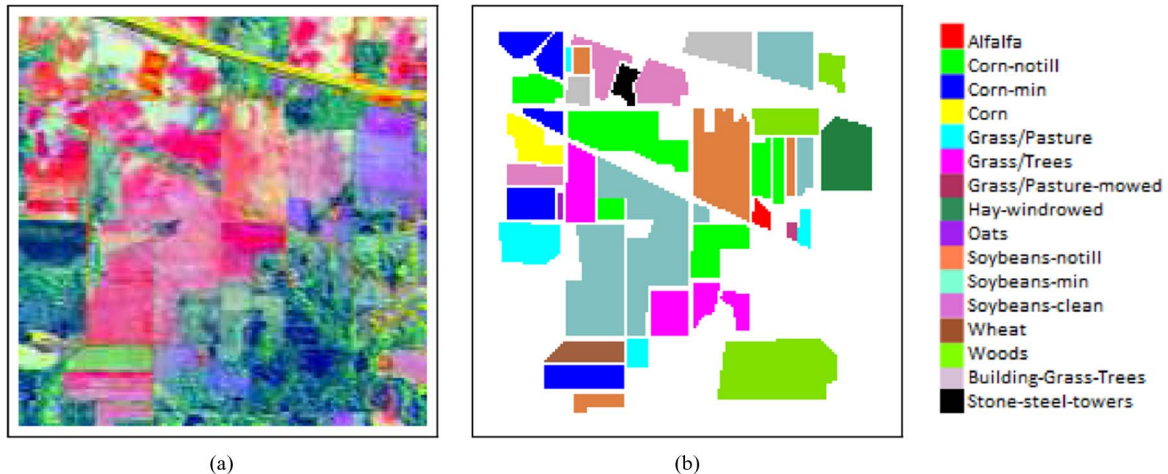
Fig. 3. (a) False color composition of the AVIRIS Indian Pines scene. (b) Ground truth map containing (right) 16 mutually exclusive land cover classes.

learning process, the fractional abundances can be precomputed in advance, and as a result, the main computational load of the proposed approach is carried out by the semisupervised classification process. According to our previous work in [25], the semisupervised learning stage is quite efficient in computational terms and able to manage a high number of unlabeled training samples without significantly increasing the computational complexity.

In the following section, we evaluate the proposed methodology using two different hyperspectral data sets that have been widely used in the hyperspectral imaging community to validate the performance of classification algorithms.

III. EXPERIMENTAL RESULT

Here, we describe the experimental results that we have conducted to validate the newly proposed technique. As mentioned in Section II-B, in this paper, we consider two widely used discriminative classifiers, namely, MLR and probabilistic SVM, due to their capacity to produce classification probabilities at a pixel level. For the MLR classifier, we optimized the parameter settings according to the procedure indicated in [16]. For the SVM classifier, we used a Gaussian radial basis function kernel and carefully optimized the kernel parameters by a grid search procedure. The optimal parameters were selected by using tenfold cross-validation. Both classifiers were carefully optimized in order to obtain the best possible performance. As it will be shown, in our experiments, the MLR classifier provided generally better results than those provided by the SVM. The spectral unmixing methodology is the one described in Section II-A. For evaluation purposes, two different data sets, which are well known in the hyperspectral imaging community, are considered. First, we study the performance of the proposed approach using a hyperspectral image with relatively low spatial resolution, collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over the Indian Pines region in Northwest Indiana. Then, we conduct experiments using a hyperspectral image with higher spatial resolution, collected by the Reflective Optics Spectrographic Imaging System (ROSIS) over the University of Pavia, Italy. All the reported values

of overall accuracy (OA), average accuracy (AA), and kappa statistic (κ) have been obtained as the average of ten Monte Carlo runs to guarantee statistical consistency.

A. Hyperspectral Data Sets

Two different real hyperspectral data sets are used in this paper. Particularly, we have used two different images, which provide different characteristics in terms of spatial and spectral resolutions in order to validate the method in very different scenarios. The scenes are collected by two different sensors: AVIRIS and ROSIS. The images span a wide range of land cover use, from agricultural areas in the Indian Pines region to urban areas in the town of Pavia, Italy. Here, we provide a description of the two aforementioned data sets.

1) AVIRIS Indian Pines: The first data set used in our experiments was collected by the AVIRIS sensor over the Indian Pines region in Northwestern Indiana in 1992. This scene, with a size of 145 lines by 145 samples, was acquired over a mixed agricultural/forest area, early in the growing season. The scene consists of 202 spectral channels in the wavelength range from 0.4 to 2.5 μm, nominal spectral resolution of 10 nm, moderate spatial resolution of 20 m by pixel, and 16-bit radiometric resolution. After an initial screening, several spectral bands were removed from the data set due to noise and water absorption phenomena, leaving a total of 164 radiance channels to be used in the experiments. For illustrative purposes, Fig. 3(a) shows a false color composition of the AVIRIS Indian Pines scene, whereas Fig. 3(b) shows the ground truth map available for the scene, displayed in the form of a class assignment for each labeled pixel, with 15 mutually exclusive ground truth classes. These data, including ground truth information, are available online,[2] a fact that has made this scene a widely used benchmark for testing the accuracy of hyperspectral data classification algorithms.

2) ROSIS Pavia University: The second data set used in experiments was collected by the ROSIS optical sensor over

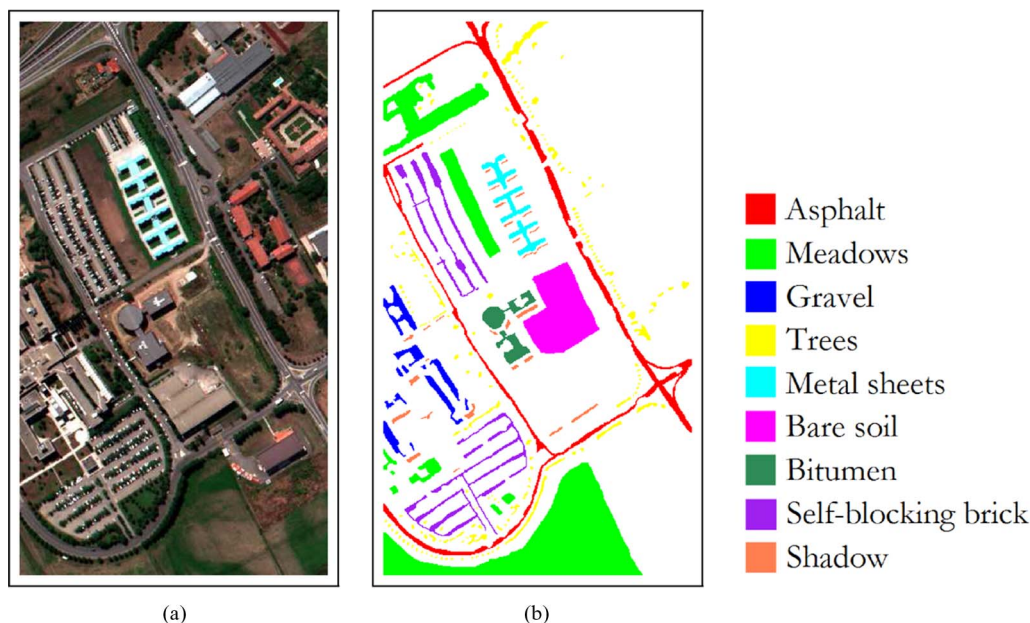[2]http://dynamo.ecn.purdue.edu/biehl/MultiSpec

Fig. 4. (a) False color composition of the ROSIS Pavia scene. (b) Ground truth map containing nine mutually exclusive classes.

the urban area of the University of Pavia, Italy. The flight was operated by Deutschen Zentrum for Luftund Raumfahrt (DLR, the German Aerospace Agency) in the framework of the Hy-Sens project, managed and sponsored by the European Union. The image size in pixels is 610×340, with very high spatial resolution of 1.3 m per pixel. The number of data channels in the acquired image is 115 (with spectral range from 0.43 to 0.86 μm). Fig. 4(a) shows a false color composite of the image, whereas Fig. 4(b) shows nine ground truth classes of interest from the ground truth.

B. Experiments With the AVIRIS Indian Pines Scene

In the first set of experiments, we use the AVIRIS Indian Pines scene in Fig. 3(a) for evaluation. Very small sets of labeled samples (5, 10, and 15 labeled samples per class) from the ground truth image in Fig. 3(b) are used as the labeled training set, and the remaining samples are used for testing. Table I reports the classification accuracies obtained by the proposed semisupervised approach in comparison with the supervised approach in [16] and the semisupervised approach in [25], using different numbers of labeled samples per class. In the semisupervised cases, 300 unlabeled samples are considered for training under three different strategies: in the first one, we selected one unlabeled sample per iteration, following the procedure described in [25]; in the second case, we select ten unlabeled samples per iteration; in the third case, we select 30 unlabeled samples per iteration. As aforementioned, the latter two strategies are intended to reduce computation time while, at the same time, evaluating the classification performance of our proposed approach using different configurations.

As shown in Table I, the results obtained by the proposed semisupervised approach are remarkable, outperforming the other tested methods in all cases. Another important observa-

tion is that, from the viewpoint of semisupervised learning, the results are always better when we include one and ten unlabeled samples per iteration, than in the case in which we included 30 samples per iteration, whereas the results obtained by using ten unlabeled samples per iteration are better or comparable with those obtained by using only one unlabeled sample per iteration. This is expected, since the AVIRIS Indian Pines data set has coarse spatial resolution of 20 m per pixel, which results in mixed pixels. In order to increase the confidence of unlabeled samples in this particular case, using only a few unlabeled samples can lead to better performance. For the case in which 30 unlabeled samples per iteration were used, in this experiment, we observed that the more unlabeled samples we include per iteration, the less confidence we have in them. As illustrated in Fig. 2, spectral unmixing can provide complementary information for the less confident samples, but a high number of less confident samples can lead to a reduction in the classification accuracy, as illustrated in Table I. In turn, by including one or ten unlabeled samples per iteration, we have higher confidence in the samples, and their associated information can be exploited more efficiently. Finally, Table I also reveals that the results obtained by using the MLR classifier are more competitive than those obtained by using the probabilistic SVM classifier in case limited training samples are available. When the number of training samples was relatively larger, for instance, 15 samples per class, both the MLR and the SVM obtained very similar results. This is consistent with the observations from our previous work [25].

For illustrative purposes, Fig. 5 shows some of the obtained classification maps by the MLR and SVM classifiers by using ten labeled samples per class. For the semisupervised cases, 300 unlabeled samples (ten per iteration) are included. It should be noted that the maps displayed in Fig. 5 correspond to one of the ten Monte Carlo runs conducted in each case. As shown in Fig. 5, the proposed semisupervised approach clearly

TABLE I

OVERALL (OA) AND AVERAGE (AA) CLASSIFICATION ACCURACIES AND κ STATISTIC OBTAINED USING DIFFERENT CLASSIFICATION STRATEGIES (BASED ON 5, 10, AND 15 LABELED SAMPLES PER CLASS) FOR THE AVIRIS INDIAN PINES HYPERSPECTRAL DATA SET WITH MLR AND SVM CLASSIFIERS. IN ALL CASES, THE RESULTS CORRESPOND TO THE MEAN VALUES OBTAINED AFTER TEN MONTE CARLO RUNS, AND THE STANDARD DEVIATIONS ARE ALSO REPORTED. THE BEST CLASSIFICATION RESULTS IN TERMS OF OA ARE DISPLAYED IN BOLD TYPEFACE

Multinomial logistic regression (MLR) classifier								
5 labeled samples per class								
OA	AA	κ	OA	AA	κ	OA	AA	κ
1 unlabeled sample per iteration			10 unlabeled samples per iteration			30 unlabeled samples per iteration		
Supervised [16]	53.58±1.83	64.46±2.45	48.43±2.09	53.58±1.83	64.46±2.45	48.43±2.09	53.58±1.83	64.46±2.45	48.43±2.09
Unmixing only	53.43±1.89	62.74±3.08	48.29±2.17	53.43±1.89	62.74±3.08	48.29±2.17	53.43±1.89	62.74±3.08	48.29±2.17
Semi-supervised [25]	66.94±4.81	70.45±3.81	62.36±5.58	70.52±3.13	76.72±1.93	66.59±3.53	68.74±3.13	76.63±2.63	64.81±3.60
Proposed approach	**74.38±3.66**	81.25±2.97	71.08±4.17	73.79±3.79	82.43±2.13	70.56±4.18	64.46±2.25	77.63±2.27	60.41±2.48
10 labeled samples per class								
OA	AA	κ	OA	AA	κ	OA	AA	κ
1 unlabeled sample per iteration			10 unlabeled samples per iteration			30 unlabeled samples per iteration		
Supervised [16]	59.45±1.79	70.77±1.73	54.62±1.79	59.45±1.79	70.77±1.73	54.62±1.79	59.45±1.79	70.77±1.73	54.62±1.79
Unmixing only	59.05±1.72	67.80±2.79	54.21±1.73	59.05±1.72	67.80±2.79	54.21±1.73	59.05±1.72	67.80±2.79	54.21±1.73
Semi-supervised [25]	67.32±6.01	75.54±2.39	63.20±6.48	73.57±1.59	80.94±1.26	72.36±1.75	73.49±3.04	81.14±1.24	70.06±3.31
Proposed approach	77.70±5.64	84.64±2.91	74.85±6.21	**81.22±2.74**	87.33±1.23	78.80±3.02	73.58±3.15	84.15±1.62	70.36±3.43
15 labeled samples per class								
OA	AA	κ	OA	AA	κ	OA	AA	κ
1 unlabeled sample per iteration			10 unlabeled samples per iteration			30 unlabeled samples per iteration		
Supervised [16]	63.63±1.53	74.33±1.48	59.21±1.71	63.63±1.53	74.33±1.48	59.21±1.71	63.63±1.53	74.33±1.48	59.21±1.71
Unmixing only	62.80±1.42	68.65±2.21	58.36±1.58	62.80±1.42	68.65±2.21	58.36±1.58	62.80±1.42	68.65±2.21	58.36±1.58
Semi-supervised [25]	71.26±4.21	79.35±2.56	67.58±4.62	77.22±1.57	82.44±1.33	74.21±1.75	76.34±1.86	83.60±1.33	73.33±2.03
Proposed approach	80.67±2.95	87.34±2.01	78.18±3.29	**83.47±1.45**	89.12±1.14	81.30±1.63	78.63±1.45	87.42±0.95	75.96±1.58
Probabilistic support vector machine (SVM) classifier								
5 labeled samples per class								
OA	AA	κ	OA	AA	κ	OA	AA	κ
1 unlabeled sample per iteration			10 unlabeled samples per iteration			30 unlabeled samples per iteration		
Supervised [16]	42.56±4.22	51.65±5.82	35.42±4.37	42.56±4.22	51.65±5.82	35.42±4.37	42.56±4.22	51.65±5.82	35.42±4.37
Unmixing only	53.43±1.89	62.74±3.08	48.29±2.17	53.43±1.89	62.74±3.08	48.29±2.17	53.43±1.89	62.74±3.08	48.29±2.17
Semi-supervised [25]	49.81±4.70	42.95±7.87	39.23±5.45	59.84±3.35	60.44±5.39	52.30±4.68	61.97±3.26	66.81±3.72	56.05±4.31
Proposed approach	55.71±7.87	59.37±11.15	46.81±10.60	**70.35±3.87**	77.33±3.86	66.42±4.40	61.98±2.36	74.39±2.56	57.56±2.36
10 labeled samples per class								
OA	AA	κ	OA	AA	κ	OA	AA	κ
1 unlabeled sample per iteration			10 unlabeled samples per iteration			30 unlabeled samples per iteration		
Supervised [16]	54.58±3.34	67.22±2.34	49.24±3.55	54.58±3.34	67.22±2.34	49.24±3.55	54.58±3.34	67.22±2.34	49.24±3.55
Unmixing only	59.05±1.72	67.80±2.79	54.21±1.73	59.05±1.72	67.80±2.79	54.21±1.73	59.05±1.72	67.80±2.79	54.21±1.73
Semi-supervised [25]	62.12±5.80	66.81±5.50	56.22±7.38	70.82±3.93	76.08±2.74	66.82±4.39	70.42±2.28	77.60±1.61	66.53±2.51
Proposed approach	75.51±6.52	82.10±6.78	71.83±8.92	**78.75±3.33**	85.80±1.16	76.00±3.65	72.14±2.73	83.18±1.35	68.70±2.93
15 labeled samples per class								
OA	AA	κ	OA	AA	κ	OA	AA	κ
1 unlabeled sample per iteration			10 unlabeled samples per iteration			30 unlabeled samples per iteration		
Supervised [16]	58.55±2.74	71.22±3.13	53.67±2.91	58.55±2.74	71.22±3.13	53.67±2.91	58.55±2.74	71.22±3.13	53.67±2.91
Unmixing only	62.80±1.42	68.65±2.21	58.36±1.58	62.80±1.42	68.65±2.21	58.36±1.58	62.80±1.42	68.65±2.21	58.36±1.58
Semi-supervised [25]	68.72±4.97	74.91±3.91	63.91±6.27	74.40±3.91	80.28±1.95	70.64±1.19	70.42±2.28	77.60±1.61	66.53±2.51
Proposed approach	82.97±5.28	88.34±3.52	80.52±6.33	**83.35±1.56**	89.62±1.34	81.16±1.74	72.14±2.73	83.18±1.35	68.70±2.93

outperforms the other two considered methods in all cases. In order to further illustrate the relative importance of discriminative classification and spectral unmixing in the conducted experiments, in Fig. 5, we also report the dominant source of information used for the classification of each pixel in the classification experiments, using a similar convention with regard to the one adopted in the toy example given in Fig. 2(e). These plots provide an indication of the importance of using

Multinomial logistic regression (MLR) classifier

| Supervised [16] (59.45%) | Unmixing only (59.05%) | Semi-supervised [25] (73.57%) | Proposed approach (81.22%) | Source of information |

Probabilistic support vector machine (SVM) classifier

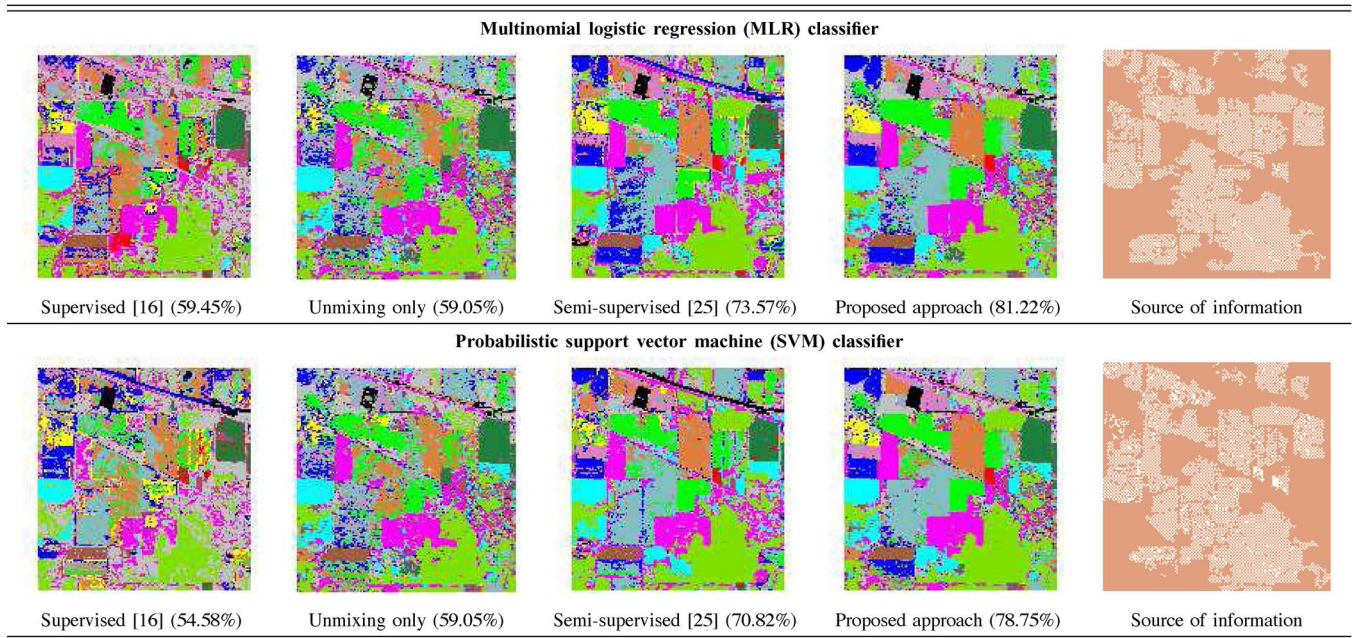| Supervised [16] (54.58%) | Unmixing only (59.05%) | Semi-supervised [25] (70.82%) | Proposed approach (78.75%) | Source of information |

Fig. 5. Classification maps along with the OAs (in parentheses) obtained using different classification strategies (based on ten labeled samples per class) for the AVIRIS Indian Pines data set with MLR and SVM classifiers. For the semisupervised cases, a total of 300 unlabeled samples (ten new samples per iteration) is used. For the proposed method (implemented with MLR and SVM), we display the dominant source of information for each pixel, where unmixing is represented in brown color, and classification is represented in white color.

discriminative classification and spectral unmixing in hyperspectral scenes with low spatial resolution. As shown in Fig. 5, the pixels dominated by classification and unmixing information tend to interact with each other and appear in the form of a chessboard distribution. This is because the pixels with more confidence in the classification (under a local and a global assumption) are those in the first-order neighborhood, whereas the other ones that do not hold this assumption require spectral unmixing as their main source of information. This is similar to the example reported in Fig. 2(c)–(e), in which the pixel at the top of the labeled sample corresponding to class number 2 in Fig. 2(a) is excluded from the set $\mathcal{X}_c$, which means that it is included in the set $\mathcal{X}_m$. Therefore, it is expected that, with more iterations, the source map turns to a chessboard distribution.

Another important observation from the maps showing the dominant source of information in Fig. 5 is that the samples that use classification as their main source of information are mainly extended from the initial labeled samples, which is consistent with the final classification maps in the sense that the regions that use classification information are generally better classified as we have more confidence in those samples. This is because the generation of the candidate set $\mathcal{D}_c$ depends on the initial training set, which is randomly selected from the ground truth image. In our experiments, we terminated the algorithm after 300 unlabeled samples were selected (as we felt that 300 iterations is a reasonable number for illustrating the performance of our proposed approach). If we performed more iterations for the semisupervised learning process, in order to enlarge the candidate set until the whole image was considered, we would end up with a full chessboard image.

C. Experiments With the ROSIS Pavia University Scene

In this second set of experiments, we use the ROSIS Pavia University data set in Fig. 4(a) to analyze the performance of the proposed approach in a challenging urban data analysis scenario with higher spatial resolution. As in the previous experiment, we considered the MLR classifier and the SVM classifier trained with {5, 10, 15} labeled samples per class from the ground truth image in Fig. 4(b). These samples were used for training, and the remaining samples were used for validation. Our semisupervised approach was again compared with the supervised method in [16] and the semisupervised method in [25]. In the semisupervised cases, we selected 300 unlabeled samples under three selection strategies: 1 unlabeled sample per iteration, 10 unlabeled samples per iteration, and 30 unlabeled samples per iteration. The latter two strategies are intended to reduce computation time while, at the same time, evaluating the classification performance of our proposed approach using different configurations. Although we use exactly the same configurations already adopted in the experiments with the AVIRIS Indian Pines scene, we would like to emphasize that different numbers of (initial) labeled samples and unlabeled samples per iteration can be considered according to the considered case study. However, we decided to use the same configuration in the experiments with both scenes since we obtained satisfactory results in both cases, which indicates that the proposed methodology does not require significant effort in terms of parameter settings.

Table II reports the obtained classification accuracies in the experiments with the ROSIS Pavia University scene. We can observe in Table II that the proposed approach obtains significant improvements over the other compared methods, in particular, when one unlabeled sample was included per active

TABLE II
OVERALL (OA) AND AVERAGE (AA) CLASSIFICATION ACCURACIES AND κ STATISTIC OBTAINED USING DIFFERENT CLASSIFICATION STRATEGIES (BASED ON 5, 10, AND 15 LABELED SAMPLES PER CLASS) FOR THE ROSIS PAVIA UNIVERSITY DATA SET WITH MLR AND SVM CLASSIFIERS. IN ALL CASES, THE RESULTS CORRESPOND TO THE MEAN VALUES OBTAINED AFTER TEN MONTE CARLO RUNS, AND THE STANDARD DEVIATIONS ARE ALSO REPORTED. THE BEST CLASSIFICATION RESULTS IN TERMS OF OA ARE DISPLAYED IN BOLD TYPEFACE

Multinomial logistic regression (MLR) classifier classifier								
	5 labeled samples per class							
	1 unlabeled sample per iteration			10 unlabeled samples per iteration			30 unlabeled samples per iteration		
	OA	AA	κ	OA	AA	κ	OA	AA	κ
Supervised [16]	64.23±3.78	72.92±1.67	55.38±3.63	64.23±3.78	72.92±1.67	55.38±3.63	64.23±3.78	72.92±1.67	55.38±3.63
Unmixing only	45.85±4.31	53.22±1.96	35.09 ±4.06	45.85±4.31	53.22±1.96	35.09±4.06	45.85± 4.31	53.22±1.96	35.09±4.06
Semi-supervised [25]	78.22±2.55	79.81±3.41	71.55±3.61	73.38±4.44	82.41±1.63	66.87±4.72	72.78±4.47	81.17±1.71	65.89±4.94
Proposed approach	**78.53±2.36**	80.11±1.83	72.53±2.78	58.35±3.77	68.59±1.8	49.54±3.95	49.57±4.1	59.1±1.76	39.46±4.03
	10 labeled samples per class								
	1 unlabeled sample per iteration			10 unlabeled samples per iteration			30 unlabeled samples per iteration		
	OA	AA	κ	OA	AA	κ	OA	AA	κ
Supervised [16]	70.14±2.5	78.91±1.78	62.64±2.67	70.14±2.5	78.91±1.78	62.64±2.67	70.14±2.5	78.91±1.78	62.64±2.67
Unmixing only	52.91±5.06	56.35±2.68	42.61 ±5.02	52.91±5.06	56.35±2.68	42.61±5.02	52.91±5.06	56.35±2.68	42.61±5.02
Semi-supervised [25]	83.72±2.37	83.62±1.14	78.56±2.87	80.12±2.78	85.58±1.19	74.65±3.19	78.73±3.38	85.53±1.78	73.12±4.01
Proposed approach	**86.97±2.92**	87.48±1.97	83.1±3.61	70.57±2.82	77.18±1.7	63.35±3.22	59.22±4.74	66.71±1.93	50.16±4.94
	15 labeled samples per class								
	1 unlabeled sample per iteration			10 unlabeled samples per iteration			30 unlabeled samples per iteration		
	OA	AA	κ	OA	AA	κ	OA	AA	κ
Supervised [16]	73.59±3.05	81.32±1.05	66.79±3.27	73.59±3.05	81.32±1.05	66.79±3.27	73.59±3.05	81.32±1.05	66.79±3.27
Unmixing only	57.61±3.11	57.55±1.02	47.53 ±3.01	57.61±3.11	57.55±1.02	47.53±3.01	57.61± 3.11	57.55±1.02	47.53±3.01
Semi-supervised [25]	85.14±2.26	84.49±1.36	80.47±2.68	82.49±3.02	86.79±0.95	77.55±3.44	82.64±3.25	87.11±0.65	77.77±3.72
Proposed approach	**89.84±2.62**	89.36±1.02	86.76±3.22	76.8±2.8	81.36±0.83	70.71±3.09	65.75±3.25	71.23±1.1	57.49±3.37
Probabilistic support vector machine (SVM) classifier								
	5 labeled samples per class								
	1 unlabeled sample per iteration			10 unlabeled samples per iteration			30 unlabeled samples per iteration		
	OA	AA	κ	OA	AA	κ	OA	AA	κ
Supervised [16]	61.49±2.61	68.92±2.76	52.27±2.94	61.49±2.61	68.92±2.76	52.27±2.94	61.49±2.61	68.92±2.76	52.27±2.94
Unmixing only	45.85±4.31	53.22±1.96	35.09 ±4.06	45.85±4.31	53.22±1.96	35.09±4.06	45.85±4.31	53.22±1.96	35.09±4.06
Semi-supervised [25]	65.99±4.33	47.65±11.34	52.30±6.11	68.82±5.94	70.22±9.02	60.36±7.29	69.34±6.41	76.61±4.88	61.71±7.37
Proposed approach	**76.45±3.79**	72.58±7.91	69.04±5.06	56.36±4.43	65.48±3.17	47.09±4.97	48.98±4.08	58.27±2.18	38.77±4.02
	10 labeled samples per class								
	1 unlabeled sample per iteration			10 unlabeled samples per iteration			30 unlabeled samples per iteration		
	OA	AA	κ	OA	AA	κ	OA	AA	κ
Supervised [16]	67.18±4.10	77.71±1.65	59.57±4.25	67.18±4.10	77.71±1.65	59.57±4.25	67.18±4.10	77.71±1.65	59.57±4.25
Unmixing only	52.91±5.06	56.35±2.68	42.61 ±5.02	52.91±5.06	56.35±2.68	42.61±5.02	52.91±5.06	56.35±2.68	42.61±5.02
Semi-supervised [25]	76.41±4.42	66.04±9.12	67.10±6.72	78.30±3.05	82.52±3.16	72.32±3.54	77.84±4.06	84.53±5.92	72.02±5.11
Proposed approach	**87.74±2.51**	86.24±2.70	83.88±3.25	70.12±2.87	76.48±2.24	62.80±3.23	59.29±5.04	66.73±1.77	50.28±5.23
	15 labeled samples per class								
	1 unlabeled sample per iteration			10 unlabeled samples per iteration			30 unlabeled samples per iteration		
	OA	AA	κ	OA	AA	κ	OA	AA	κ
Supervised [16]	73.47±5.19	81.64±1.64	66.85±5.67	73.47±5.19	81.64±1.64	66.85±5.67	73.47±5.19	81.64±1.64	66.85±5.67
Unmixing only	57.61±3.11	57.55±1.02	47.53 ±3.01	57.61±3.11	57.55±1.02	47.53±3.01	57.61± 3.11	57.55±1.02	47.53±3.01
Semi-supervised [25]	78.21±7.91	76.44±6.05	71.43±8.83	82.55±2.88	86.19±1.04	77.61±3.26	80.32±5.61	86.39±1.97	75.08±6.41
Proposed approach	**89.15±6.22**	89.56±3.64	86.02±7.57	77.46±2.97	81.87±1.13	71.54±3.32	65.90±3.72	71.57±1.2	57.71±3.91

learning iteration. However, it can be also noticed that the performance of the proposed approach decreases significantly when the number of unlabeled samples per iteration is in-creased. This is in contrast with the experiments reported for the AVIRIS Indian Pines scene. This is due to the different spatial resolutions of the considered scenes. For instance, in the
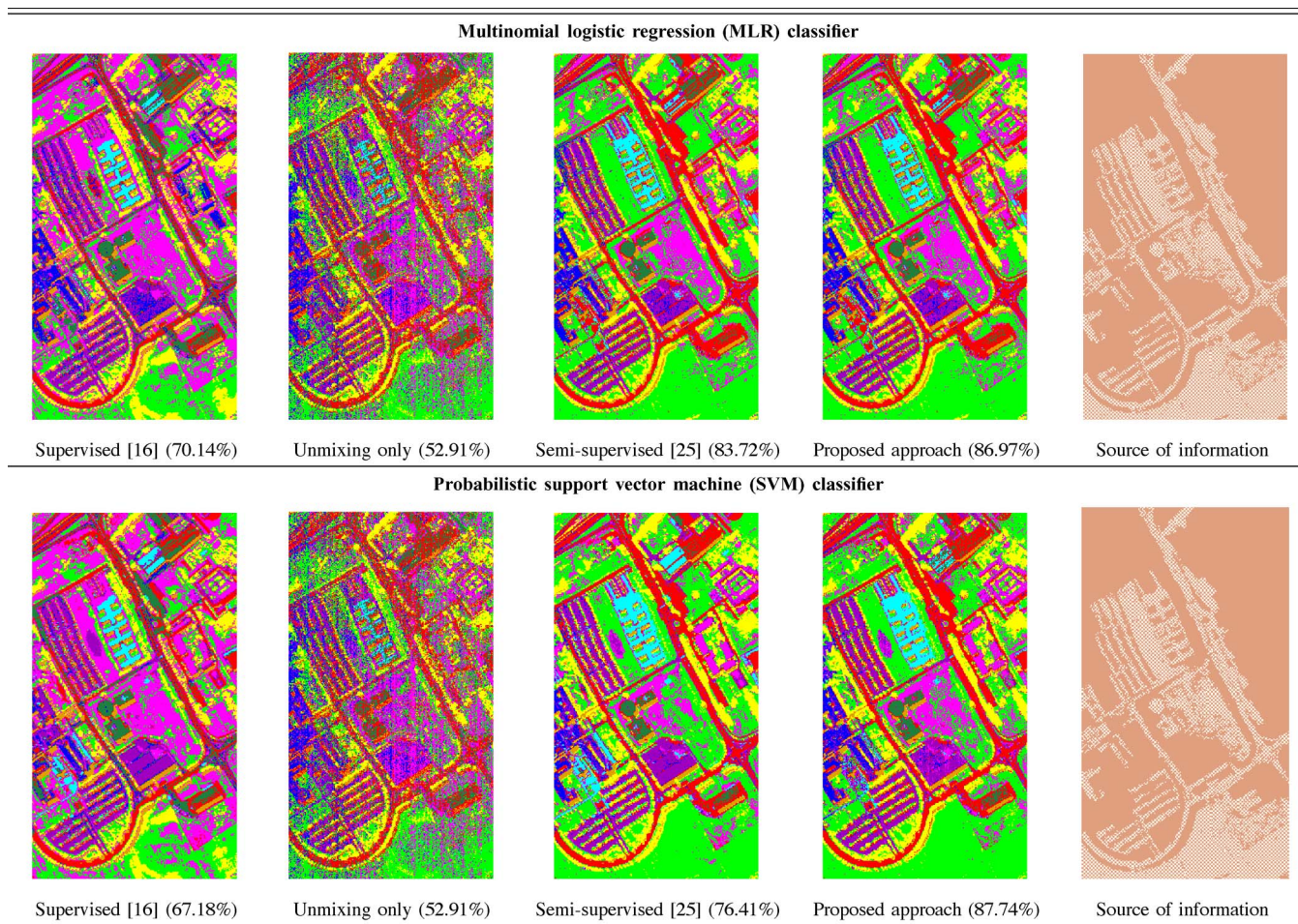
Multinomial logistic regression (MLR) classifier

Supervised [16] (70.14%) Unmixing only (52.91%) Semi-supervised [25] (83.72%) Proposed approach (86.97%) Source of information

Probabilistic support vector machine (SVM) classifier

Supervised [16] (67.18%) Unmixing only (52.91%) Semi-supervised [25] (76.41%) Proposed approach (87.74%) Source of information

Fig. 6.   Classification maps along with the OAs (in parentheses) obtained using different classification strategies (based on ten labeled samples per class) for the ROSIS Pavia University data set with MLR and SVM classifiers. For the semisupervised cases, a total of 300 unlabeled samples (1 new sample per iteration) is used. For the proposed method (implemented with MLR and SVM), we display the dominant source of information for each pixel, where unmixing is represented in brown color, and classification is represented in white color.

AVIRIS Indian Pines scene, the agricultural features were very early in their growth cycle at the time of data collection. As a result, the linear mixture model can explain most of the pixels in the scene, and there is no risk in including additional (less confident) unlabeled samples per iteration since it is expected that spectral unmixing can properly characterize those samples. However, in the ROSIS Pavia University scene (despite its higher spatial resolution), the complexity of the mixtures is expected to be higher due to the complex urban landscape. As a result, the confidence in the selected unlabeled samples is also important. This is achieved when the number of unlabeled samples selected per iteration is decreased, as more confidence is assigned to these samples in the discriminative classification process. A final observation is that, in comparison with the SVM classifier, the advantages of the MLR classifier decrease as the number of training samples becomes larger.

For illustrative purposes, Fig. 6 shows some of the classification maps obtained by the MLR and probabilistic SVM classifiers, respectively, for the ROSIS Pavia University scene, using 10 labeled samples and 300 unlabeled samples (one per iteration) in the semisupervised cases. These classification maps correspond to one of the ten Monte Carlo runs conducted for

each experiment. The improvements obtained by the proposed strategy can be clearly appreciated. These maps reveal that, even in a hyperspectral image with high spatial resolution, significant improvements can still be obtained by integrating the complementary nature of classification and spectral unmixing in the analysis process.

As in the case with the AVIRIS Indian Pines experiments, Fig. 6 also reports the dominant source of information (classification or unmixing) used for each pixel, using a similar convention with regard to the one adopted in the toy example given in Fig. 2(e). These plots provide an indication of the importance of using discriminative classification and spectral unmixing in hyperspectral scenes with high spatial resolution, in which mixed pixels also need to be characterized as indicated by our experiments. Similar conclusions with regard to the AVIRIS Indian Pines experiments can be drawn, and a general conclusion is that the proposed approach can properly take advantage of the complementarity of spectral unmixing and discriminative classification. This is an important observation, as these two sources of information have been rarely exploited in complementary fashion in the past.

IV. CONCLUSION AND FUTURE WORK

In this paper, we have developed a new framework for the complementary integration of two sources of information, i.e., discriminative classification and spectral unmixing, in the analysis of remotely sensed hyperspectral data sets. A main innovation of the proposed approach is that it adaptively uses the information provided by probabilistic classifiers and spectral unmixing in a semisupervised classification process, where the samples with high confidence (most likely pure in spectral terms) are characterized by a discriminative classifier, and the samples with less confidence (probably mixed in nature) are characterized by a standard spectral unmixing chain. As a result, the proposed approach naturally integrates the information provided by discriminative classification and spectral unmixing adaptively and does not require any weight parameters while exploiting both classification and unmixing in synergetic fashion. Our experimental results, conducted using well-known hyperspectral scenes with different spatial resolutions, indicate that spectral unmixing and discriminative classification exhibit great complementarity, which can be exploited (in semisupervised fashion) to obtain competitive classification results with regard to other similar approaches recently presented in the literature.

In future work, we are planning on developing strategies to further reduce the computational complexity of the presented approaches, although the current complexity is not significantly higher than the one exhibited by the traditional semisupervised classification process. Specifically, our approach benefits from the fact that the spectral unmixing information can be computed a priori and quite effectively; hence, the complexity of our approach is driven by semisupervised classification. In future work, we are also planning on developing high-performance computing implementations that can take advantage of the inherent parallel nature of some of the steps of the presented methodology, to be able to implement the proposed approach for the whole image samples (in the current experiments, we only exploited 300 iterations for practical reasons, while a deeper exploration of the whole scene would be desirable). Additional experiments with other scenes with different spatial and spectral resolutions, also depicting different environments, are also desirable in future developments of the method. In the future, we will use soft classification methods instead of spectral unmixing methods in order to exploit soft classification labels in combination with hard classification labels, as suggested in [49].

ACKNOWLEDGMENT

REFERENCES

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

[2] A. Plaza, J. M. Bioucas-Dias, A. Simic, and W. J. Blackwell, "Foreword to the special issue on hyperspectral image and signal processing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 347–353, Apr. 2012.

[3] Q. Du *et al.*, "Foreword to the special issue on hyperspectral remote sensing: Theory, methods, and applications," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 459–465, Apr. 2013.

[4] 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.

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

[6] N. Younan, S. Aksoy, and R. King, "Foreword to the special issue on pattern recognition in remote sensing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 5, pp. 1331–1334, Oct. 2012.

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

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

[9] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis: An Introduction*. New York, NY, USA: Springer-Verlag, 2006.

[10] J. Borges, Bioucas-Dias, and A. Marçal, "Bayesian hyperspectral image segmentation with discriminative class learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2151–2164, Jun. 2011.

[11] A. Castrodad *et al.*, "Learning discriminative sparse representations for modeling, source separation, and mapping of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4263–4281, Nov. 2011.

[12] L. Zhang, L. Zhang, D. Tao, and X. Huang, "Tensor discriminative locality alignment for hyperspectral image spectral–spatial feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 242–256, Jan. 2013.

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

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

[15] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 809–823, Mar. 2012.

[16] J. Li, J. Bioucas-Dias, and A. Plaza, "Hyperspectral image segmentation using a new Bayesian approach with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3947–3960, Oct. 2011.

[17] Y. Chen, N. Nasrabadi, and T. Tran, "Hyperspectral image classification via kernel sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 217–231, Jan. 2013.

[18] M. Chi and L. Bruzzone, "Semisupervised classification of hyperspectral images by SVMs optimized in the primal," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1870–1880, Jun. 2007.

[19] L. Bruzzone, M. Chi, and M. Marconcini, "A novel transductive SVM for the semisupervised classification of remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3363–3373, Nov. 2006.

[20] G. Camps-Valls, T. Bandos, and D. Zhou, "Semi-supervised graph-based hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3044–3054, Oct. 2007.

[21] D. Tuia and G. Camps-Valls, "Semisupervised remote sensing image classification with cluster kernels," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 2, pp. 224–228, Apr. 2009.

[22] D. Tuia and G. Camps-Valls, "Urban image classification with semisupervised multiscale cluster kernels," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 1, pp. 65–74, Mar. 2011.

[23] J. Li, J. Bioucas-Dias, and A. Plaza, "Semi-supervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4085–4098, Nov. 2010.

[24] L. Bruzzone and C. Persello, "A novel context-sensitive semisupervised SVM classifier robust to mislabeled training samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2142–2154, Jul. 2009.

[25] I. Dopido *et al.*, "Semi-supervised self-learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 4032–4044, Jul. 2013.

[26] M. Crawford, D. Tuia, and H. Yang, "Active learning: Any value for classification of remotely sensed data?" *Proc. IEEE*, vol. 101, no. 3, pp. 593–608, Mar. 2013.

[27] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. 110–122, Sep. 2009.

[28] W.-K. Ma *et al.*, "A signal processing perspective on hyperspectral unmixing: Insights from remote sensing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 67–81, Sep. 2014.

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

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

[31] Q. Du, N. Raksuntorn, N. H. Younan, and R. L. King, "End-member extraction for hyperspectral image analysis," *Appl. Opt.*, vol. 47, no. 28, pp. 77–84, Oct. 2008.

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

[33] D. Anderson and A. Zare, "Spectral unmixing cluster validity index for multiple sets of endmembers," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 4, pp. 1282–1295, Aug. 2012.

[34] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.

[35] I. Dopido, M. Zortea, A. Villa, A. Plaza, and P. Gamba, "Unmixing prior to supervised classification of remotely sensed hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 6, pp. 760–764, Jul. 2011.

[36] M. Xu, B. Du, and L. Zhang, "Spatial–spectral information based abundance-constrained endmember extraction methods," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2004–2015, Jun. 2014.

[37] I. Dopido, A. Villa, A. Plaza, and P. Gamba, "A quantitative and comparative assessment of unmixing-based feature extraction techniques for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 421–435, Apr. 2012.

[38] I. Dopido, J. Li, P. Gamba, and A. Plaza, "A new hybrid strategy combining semi-supervised classification and unmixing of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 8, Aug. 2014, Art. ID 2322143.

[39] D. Heinz and C.-I. Chang, "Fully constrained least squares linear mixture analysis for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 3, pp. 529–545, Mar. 2001.

[40] J. M. P. Nascimento and J. M. Bioucas-Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

[41] J. Boardman, "Leveraging the high dimensionality of AVIRIS data for improved subpixel target unmixing and rejection of false positives: Mixture tuned matched filtering," *Proc. 5th JPL Geosci. Workshop*, 1998, pp. 55–56.

[42] D. Böhning, "Multinomial logistic regression algorithm," *Ann. Inst. Statist. Math.*, vol. 44, no. 1, pp. 197–200, Mar. 1992.

[43] B. Krishnapuram, L. Carin, M. Figueiredo, and A. Hartemink, "Sparse multinomial logistic regression: Fast algorithms and generalization bounds," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 6, pp. 957–968, Jun. 2005.

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

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

[46] J. Platt, "Probabilities for support vector machines," in *Advances in Large Margin Classifiers*. Cambridge, MA, USA: MIT Press, 2000, pp. 61–74.

[47] D. Tuia, M. Volpi, L. Copa, M. Kanevski, and J. Munoz-Mari, "A survey of active learning algorithms for supervised remote sensing image classification," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 606–617, Jun. 2011.

[48] T. Luo *et al.*, "Active learning to recognize multiple types of plankton," *J. Mach. Learn. Res.*, vol. 6, pp. 589–613, 2005.

[49] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semi-supervised hyperspectral image classification using soft sparse multinomial logistic regression," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 2, pp. 318–322, Mar. 2013.

Jun Li (M'13) received the B.S. degree in geographic information systems from Hunan Normal University, Changsha, China, in 2004, the M.E. degree in remote sensing from Peking University, Beijing, China, in 2007, and the Ph.D. degree in electrical engineering from Universidade Técnica de Lisboa, Lisbon, Portugal, in 2011.

From 2007 to 2011, she was a Marie Curie Research Fellow with the Departamento de Engenharia Electrotécnica e de Computadores and the Instituto de Telecomunicações, Instituto Superior Técnico, Universidade Técnica de Lisboa, in the framework of the European Doctorate for Signal Processing (SIGNAL). She has been also actively involved in the Hyperspectral Imaging Network, a Marie Curie Research Training Network involving 15 partners in 12 countries and intended to foster research, training, and cooperation on hyperspectral imaging at the European level. Since 2011, she has been a Postdoctoral Researcher with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, Cáceres, Spain. Her research interests include hyperspectral image classification and segmentation, spectral unmixing, signal processing, and remote sensing.

Dr. Li has been a reviewer of several journals, including the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, *Pattern Recognition, Optical Engineering, Journal of Applied Remote Sensing*, and *Inverse Problems and Imaging*. She was a recipient the Best Reviewer Award of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING in 2012.

Inmaculada Dópido received the B.S. and M.S. degrees in telecommunications and the M.S. and Ph.D. degrees in computer engineering from the University of Extremadura, Caceres, Spain.

She has been a Member of the Hyperspectral Computing Laboratory coordinated by Prof. Antonio Plaza. She has been a Visiting Researcher with the Department of Electronics, University of Pavia, Pavia, Italy, and École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. She is currently a Postdoctoral Researcher with the Institute for Applied Microelectronics, University of Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain. Her research interests include remotely sensed hyperspectral imaging, pattern recognition and signal and image processing, with a particular emphasis on the development of new techniques for unsupervised and supervised classifications, and spectral mixture analysis of hyperspectral data.

Dr. Dópido is a Reviewer for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, the IEEE TRANSACTION ON GEOSCIENCE AND REMOTE SENSING, and the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.

Paolo Gamba (SM'00–F'13) received the Laurea (*cum laude*) and Ph.D. degrees in electronic engineering from the University of Pavia, Pavia, Italy, in 1989 and 1993, respectively.

He is currently an Associate Professor of telecommunications with the University of Pavia, where he also leads the Telecommunications and Remote Sensing Laboratory. He has been invited to give keynote lectures and tutorials in several occasions about urban remote sensing, data fusion, Earth observation data, and risk management. He has authored or coauthored over 100 papers in international peer-reviewed journals and presented over 250 research works in workshops and conferences.

Dr. Gamba served as a Chair of the Data Fusion Committee of the IEEE Geoscience and Remote Sensing Society from October 2005 to May 2009 and is currently the Chair of the Chapters' committee. He has been the Organizer and the Technical Chair of the biennial GRSS/ISPRS joint workshops on "Remote Sensing and Data Fusion Over Urban Areas" since 2001. The latest edition, JURSE 2015, will be held in Lausanne, Switzerland, in April 2015. He also served as a Technical Cochair for the 2010 IEEE Geoscience and Remote Sensing Symposium held in Honolulu, Hawaii, in July 2010. He will serve as a Technical Cochair for the 2015 IEEE Geoscience and Remote Sensing Symposium to be held in Milan, Italy. He served as the Editor-in-Chief of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS from 2009 to 2013. He has been a Guest Editor of special issues of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE JOURNAL OF SELECTED TOPICS IN REMOTE SENSING APPLICATIONS, the *ISPRS Journal of Photogrammetry and Remote Sensing*, the international journal *Information Fusion*, and *Pattern Recognition Letters* on the topics of Urban Remote Sensing, Remote Sensing for Disaster Management, and Pattern Recognition in Remote Sensing Applications.

Antonio Plaza (SM'05) was born in Caceres, Spain, in 1975. He received the B.S., M.Sc., and Ph.D. degrees in computer enginnering from the University of Extremadura, Caceres, Spain, in 1997, 1999, and 2002, respectively.

He is currently an Associate Professor (with accreditation for Full Professor) with the Department of Technology of Computers and Communications, University of Extremadura, Caceres, where he is also the Head of the Hyperspectral Computing Laboratory. He has been the advisor of 12 Ph.D. dissertations and over 30 M.Sc. dissertations. He was the Coordinator of the Hyperspectral Imaging Network, a European project with a total funding of 2.8 million euro. He has authored over 400 publications, including 126 Journal Citation Reports journal papers (78 in IEEE journals), 20 book chapters, and over 240 peer-reviewed conference proceeding papers (94 in IEEE conferences). He has edited a book on *High-Performance Computing in Remote Sensing* (CRC Press/Taylor and Francis, 2007) (the first book on this topic in the published literature). His main research interests include remotely sensed hyperspectral image analysis and efficient implementations of large-scale scientific problems on high-performance computing architectures.

Dr. Plaza served as the Director of Education Activities for the IEEE Geoscience and Remote Sensing Society in 2011–2012 and has been serving as the President of the society's Spanish Chapter since November 2012. He has also served as a Proposal Evaluator for the European Commission (Marie Curie Actions, Engineering Panel), the European Space Agency, the Belgium Science Policy, the Israel Science Foundation, and the Spanish Ministry of Science and Innovation. He has participated in the Tenure Track Selection Committee of different universities in Italy, Spain, and Australia. He is an Associate Editor of IEEE ACCESS and was a member of the Editorial Board of the IEEE GEOSCIENCE AND REMOTE SENSING NEWSLETTER (2011–2012) and the IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE (2013). He was also a member of the steering committee of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS). He has been a Guest Editor of eight special issues on hyperspectral remote sensing for different journals. He is currently an Associate Editor of the *Journal of Real-Time Image Processing*. He has reviewed over 500 manuscripts for over 50 different journals. He is also currently serving as the Editor-in-Chief of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He was a recipient of the Best Ph.D. Dissertation Award at the University of Extremadura in 2002. He was also a recipient of the Best Paper Award at the IEEE Symposium on Signal Processing and Information Technology in 2008. He was also a recipient of the recognition of Best Reviewers of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS in 2009 and the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING in 2010, a journal for which he served as an Associate Editor in 2007–2012. He was a coauthor of the Best Student Paper at the IEEE International Conference on Space Technology in 2011. He was also a recipient of the Best Paper Award from JSTARS in 2013 and of the most highly cited paper (2005–2010) in the Journal of Parallel and Distributed Computing.