A Quantitative and Comparative Assessment of Unmixing-Based Feature Extraction Techniques for Hyperspectral Image Classification

Inmaculada Dópido, Alberto Villa, Member, IEEE, Antonio Plaza, Senior Member, IEEE, and Paolo Gamba, Senior Member, IEEE

Abstract—Over the last years, many feature extraction techniques have been integrated in processing chains intended for hyperspectral image classification. In the context of supervised classification, it has been shown that the good generalization capability of machine learning techniques such as the support vector machine (SVM) can still be enhanced by an adequate extraction of features prior to classification, thus mitigating the curse of dimensionality introduced by the Hughes effect. Recently, a new strategy for feature extraction prior to classification based on spectral unmixing concepts has been introduced. This strategy has shown success when the spatial resolution of the hyperspectral image is not enough to separate different spectral constituents at a sub-pixel level. Another advantage over statistical transformations such as principal component analysis (PCA) or the minimum noise fraction (MNF) is that unmixing-based features are physically meaningful since they can be interpreted as the abundance of spectral constituents. In turn, previously developed unmixing-based feature extraction chains do not include spatial information. In this paper, two new contributions are proposed. First, we develop a new unmixing-based feature extraction technique which integrates the spatial and the spectral information using a combination of unsupervised clustering and partial spectral unmixing. Second, we conduct a quantitative and comparative assessment of unmixing-based versus traditional (supervised and unsupervised) feature extraction techniques in the context of hyperspectral image classification. Our study, conducted using a variety of hyperspectral scenes collected by different instruments, provides practical observations regarding the utility and type of feature extraction techniques needed for different classification scenarios.

Index Terms—Hyperspectral image classification, spatial-spectral integration, spectral unmixing, support vector machines (SVMs), unmixing-based feature extraction.

I. INTRODUCTION

The rich spectral information available in remotely sensed hyperspectral images allows for the possibility to distinguish between spectrally similar materials [1]. However, supervised classification of hyperspectral images is a very challenging task due to the generally unfavorable ratio between the (large) number of spectral bands and the (limited) number of training samples available a priori, which results in the Hughes phenomenon [2]. As shown in [3], when the number of features considered for classification is larger than a threshold, the classification accuracy starts to decrease. The application of methods originally developed for the classification of lower dimensional data sets (such as multispectral images) provides therefore poor results when applied to hyperspectral images, especially in the case of small training sets [4]. On the other hand, the collection of reliable training samples is very expensive in terms of time and finance, and the possibility to exploit large ground truth information is not common [5]. To address this issue, a dimensionality reduction step is often performed prior to the classification process, in order to bring the information in the original space (which in the case of hyperspectral data is almost empty [4]) to the right subspace which allows separating the classes by discarding information that is useless for classification purposes. Several feature extraction techniques have been proposed to reduce the dimensionality of the data prior to classification, thus mitigating the Hughes phenomenon. These methods can be unsupervised (if no a priori information is available) or supervised (if available training samples are used to project the data onto a classification-optimized subspace [6], [7]). Classic unsupervised techniques include principal component analysis (PCA) [8], the minimum noise fraction (MNF) [9], or independent component analysis (ICA) [10]. Supervised approaches comprise discriminant analysis for feature extraction (DAFE), decision boundary feature extraction (DBFE), and non-parametric weighted feature extraction (NWFE), among many others [4], [11].

In the context of supervised classification, kernel methods have been widely used due to their insensitivity to the curse of dimensionality [12]. However, the good generalization capability of machine learning techniques such as the support vector
machine (SVM) [13] can still be enhanced by an adequate extraction of relevant features to be used for classification purposes [14], especially if limited training sets are available a priori. Recently, we have investigated this issue by developing a new set of feature extraction techniques based on spectral unmixing concepts [15]. These techniques are intended to take advantage of spectral unmixing models [16] in the characterization of training samples, thus including additional information about sub-pixel composition that can be exploited at the classification stage. Another advantage of unmixing-based techniques over statistical transformations such as PCA, MNF or ICA is the fact that the features derived by spectral unmixing are physically meaningful since they can be interpreted as the abundance of spectrally pure constituents. Although unmixing-based feature extraction offers an interesting alternative to classic (supervised and unsupervised approaches), several important aspects deserve further attention [17]:

1) First, the unmixing-based chains discussed in [15] do not include spatial information, which is an important source of information since hyperspectral images exhibit spatial correlation between image features.

2) Second, the study in [15] suggested that partial unmixing [18], [19] could be an effective solution to deal with the likely fact that not all pure spectral constituents in the scene (needed for spectral unmixing purposes) are known a priori, but a more exhaustive investigation of partial unmixing (particularly in combination with spatial information) is needed.

3) Finally, the number of features to be extracted prior to classification was set in [15] to an empirical value given by the intrinsic dimensionality of the input data. However, in the context of supervised feature extraction the number of features to be retained is probably linked to the characteristics of the training set rather than the full hyperspectral image. Hence, a detailed investigation of the optimal number of features that need to be extracted prior to classification is highly desirable.

In this paper, we address the aforementioned issues by means of two highly innovative contributions. First, a new feature extraction technique exploiting sub-pixel information is proposed. This approach integrates spatial and spectral information using unsupervised clustering in order to define spatially homogeneous regions prior to the partial unmixing stage. A second contribution of this work is a detailed investigation on the issue of how many (and what type of) features should be extracted prior to SVM-based classification of hyperspectral data. For this purpose, different types of (classic and unmixing-based) feature extraction strategies, both unsupervised and supervised in nature, are considered.

The remainder of the paper is organized as follows. Section II describes a new unmixing-based feature extraction technique which integrates the spatial and the spectral information. A supervised and an unsupervised version of this technique are developed. Section III describes several representative hyperspectral scenes which have been used in our experiments. This includes three scenes collected by the Airborne Visible Infra-Red Imaging Spectrometer (AVIRIS) [20] system over the regions of Indian Pines, Indiana, Kennedy Space Center, Florida, and Salinas Valley, California, and also a hyperspectral scene collected by the Reflective Optics Spectrographic Imaging System (ROSIS) [21] over the city of Pavia, Italy. Section IV provides an experimental comparison of the proposed feature extraction chains with regards to other classic and unmixing-based approaches, using the four considered hyperspectral image scenes. Section V concludes the paper with some remarks and hints at plausible future research lines.

II. A NEW UNMIXING-BASED FEATURE EXTRACTION TECHNIQUE

This section is organized as follows. In Section II-A we fix notation and describe some general concepts about linear spectral unmixing, adopted as our baseline mixture model due to its simplicity and computational tractability. Section II-B describes an unsupervised feature extraction strategy based on spectral unmixing concepts. This strategy first performs $k$-means clustering, searching for as many classes as the number of features that need to be retained. The centroids of each cluster are considered as the endmembers, and then the features are obtained by applying spectral unmixing for abundance estimation. The main objective of this chain is to solve problems highlighted by endmember extraction based algorithms, which are sensitive to outliers and pixels with extreme values of reflectance. By using an unsupervised clustering method, the endmembers extracted are expected to be more spatially significant. Finally, Section II-C describes a modified version of the feature extraction technique in which the endmembers are searched in the available training set instead of the entire original image. Here, our assumption is that training samples may better represent the available land cover classes in the subsequent classification process.

A. Linear Spectral Unmixing

Let us denote a remotely sensed hyperspectral scene with $n$ bands by $I$, in which the pixel at the discrete spatial coordinates $(i,j)$ of the scene is represented by a vector $X(i,j) = [x_1(i,j), x_2(i,j), \ldots, x_n(i,j)] \in \mathbb{R}^n$, where $\mathbb{R}$ denotes the set of real numbers in which the pixel's spectral response $x_k(i,j)$ at sensor channels $k = 1, \ldots, n$ is included. Under the linear mixture model assumption, each pixel vector in the original scene can be modeled using the following expression:

$$X(i,j) = \sum_{z=1}^{p} \Phi_z(i,j) \cdot E_z + n(i,j)$$

(1)

where $E_z$ denotes the spectral response of endmember $z$, $\Phi_z(i,j)$ is a scalar value designating the fractional abundance of the endmember $z$ at the pixel $X(i,j)$, $p$ is the total number of endmembers, and $n(i,j)$ is a noise vector. An unconstrained solution to (1) is simply given by the following expression [22]:

$$\hat{\Phi}_{UC}(i,j) = (E^T E)^{-1} E^T X(i,j).$$

(2)

Two physical constraints are generally imposed into the model described in (1), these are the abundance non-negativity constraint (ANC), i.e., $\Phi_z(i,j) \geq 0$, and the abundance sum-to-one constraint (ASC), i.e., $\sum_{z=1}^{p} \Phi_z(i,j) = 1$ [23]. Imposing the ASC constraint results in the following optimization problem:
Similarly, imposing the ANC constraint results in the following optimization problem:

\[
\min_{\Phi(i,j) \in \Delta} \left\{ \left( \left( \mathbf{X}(i,j) - \Phi(i,j) \cdot \mathbf{E} \right) \right)^T \left( \mathbf{X}(i,j) - \Phi(i,j) \cdot \mathbf{E} \right) \right\},
\]
subject to: \( \Delta = \left\{ \Phi(i,j) \left| \sum_{z=1}^{p} \Phi_z(i,j) = 1 \right. \right\} \). (3)

As indicated in [23], a fully constrained (i.e. ASC-constrained and ANC-constrained) estimate can be obtained in least-squares sense by solving the optimization problems in (3) and (4) simultaneously. However, in order for such estimate to be meaningful, it is required that the spectral signatures of all endmembers, i.e., \( \{ \mathbf{E}_z \}_{z=1}^{p} \), are available \textit{a priori}, which is not always possible. Such fully constrained linear spectral unmixing estimate is generally referred to in the literature by the acronym FCLSU. In the case where not all endmember signatures are available in advance, partial unmixing has emerged as a suitable alternative to solve the linear spectral unmixing problem [19].

B. Unsupervised Unmixing-Based Feature Extraction

In this subsection we describe our first approach to design a new unmixing-based feature extraction technique which integrates spatial and spectral information. It can be summarized by the flowchart in Fig. 1. First, we apply the \( k \)-means algorithm [24] to the original hyperspectral image. Its goal is to determine a set of \( c \) points, called centers, so as to minimize the mean squared distance from each pixel vector to its nearest center. The algorithm is based on the observation that the optimal placement of a center is at the centroid of the associated cluster. It starts with a random initial placement. At each stage, the algorithm moves every center point to the centroid of the set of pixel vectors for which the center is a nearest neighbor according to the spectral angle (SA) [16], and then updates the neighborhood by recomputing the SA from each pixel vector to its nearest center. These steps are repeated until the algorithm converges to a point that is a minimum for the distortion [24]. The output of \( k \)-means is a set of spectral clusters, each made up of one or more spatially connected regions. In order to determine the number of clusters (endmembers) in advance, techniques used to estimate the number of endmembers like the virtual dimensionality (VD) [25] or the hyperspectral subspace identification by minimum error (HySime) [26] can be used. In our experiments we vary the number of clusters in a certain range in order to analyze the impact of this parameter. In fact, our main motivation for using a partial unmixing technique at this point is the fact that the estimation of the number of endmembers in the original image is a very challenging issue. It is possible that the actual number of endmembers in the original image, \( p \), is larger than the number of clusters derived by \( k \)-means. In this case, in order to unmix the original image we need to address a situation in which not all endmembers may be available \textit{a priori}). It has been shown in previous work that the FCLSU technique does not provide accurate results in this scenario [15]. In turn, it is also possible that \( p \leq c \). In this case, partial unmixing has shown great success [19] in abundance estimation. Following this line of reasoning, we have decided to resort to partial unmixing techniques in this work.

A successful technique to estimate abundance fractions in such partial unmixing scenarios is mixture-tuned matched filtering (MTMF) [19]—also known in the literature as constrained energy minimization (CEM) [18], [22]—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 just one endmember spectrum (that of the target of interest) and therefore behaves as a partial unmixing method that suppresses background noise.
Fig. 2. Block diagram illustrating a supervised clustering followed by MTMF \((\text{C} \text{MTMF}_{\text{sup}})\) technique for unmixing-based feature extraction.

and estimates the sub-pixel abundance of a single endmember material without assuming the presence of all endmembers in the scene, as it is the case with FCLSU. If we assume that \(E_z\) is the endmember to be characterized, MTMF estimates the abundance fraction \(\Phi(i, j)\) of \(E_z\) in a specific pixel vector \(X(i, j)\) of the scene as follows:

\[
\Phi_{\text{MTMF}}(i, j) = \left( (E_z^T R^{-1} E_z)^{-1} R^{-1} E_z \right)^T X(i, j)
\]  
(5)

where \(R\) is the matrix:

\[
R = \frac{1}{s \times l} \sum_{i=1}^{s} \sum_{j=1}^{l} X(i, j) X^T(i, j)
\]  
(6)

with \(s\) and \(l\) respectively denoting the number of samples and the number of lines in the original hyperspectral image. As shown by Fig. 1, the features resulting from the proposed unmixing-based technique, referred to hereafter as unsupervised clustering followed by MTMF \((\text{C} \text{MTMF}_{\text{unsup}})\), are used to train an SVM classifier with a few randomly selected labeled samples. The classifier is then tested using the remaining labeled samples.

C. Supervised Unmixing-Based Feature Extraction

Fig. 2 describes a variation of the \(\text{C} \text{MTMF}_{\text{unsup}}\) technique presented in the previous subsection in which the endmembers are extracted from the available (labeled) training samples instead of from the original image. This introduces two main properties with regards to \(\text{C} \text{MTMF}_{\text{unsup}}\): 1) the number of endmembers to be extracted is given by the total number of different classes, \(c\), in the labeled samples available in the training set, and 2) the endmembers (class centers) are obtained after clustering the training set, which reduces computational complexity significantly. The increase in computational performance comes at the expense of introducing an additional consideration. In this scenario, it is likely that the actual number of endmembers in the original image, \(p\), is larger than the number of different classes comprised by available labeled training samples, \(c\). Therefore, in order to unmix the original image we again need to address a partial unmixing problem. Then, as shown by Fig. 2, standard SVM classification is performed on the stack of abundance fractions using randomly selected training samples. Hereinafter, we refer to the feature extraction technique described in Fig. 2 as supervised clustering followed by MTMF \((\text{C} \text{MTMF}_{\text{sup}})\).

III. HYPSPECTRAL DATA SETS

In order to have a fair experimental comparison between the proposed and available feature extraction approaches, several representative hyperspectral data sets are investigated. In this work, we have considered four different images captured by two different sensors: AVIRIS and ROSIS. The images span a wide range of land cover use, from agricultural areas of Indian Pines and Salinas, to urban zones in the town of Pavia and mixed vegetation/urban features in Kennedy Space Center. The number of ground-truth pixels per class for all the considered hyperspectral images is given in Table I. In the following, we briefly describe each of the data sets considered in our study.

A. 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 comprises 202 spectral channels in the wavelength range from 0.4 to 2.5 \(\mu\)m, nominal spectral resolution of 10 nm, moderate spatial resolution of 20 meters 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, while 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 16 mutually exclusive ground-truth classes. These data, including ground-truth information, are available online,\(^1\) a fact which has made this scene a widely used benchmark for testing the accuracy of hyperspectral data classification algorithms.

\(^1\)http://dynamo.ecn.purdue.edu/biehl/MultiSpec.
TABLE I

NUMBER OF PIXELS IN EACH GROUND-TRUTH CLASS IN THE FOUR CONSIDERED HYPERSONTRICAL IMAGES. THE NUMBER OF TRAINING AND TEST PIXELS USED IN OUR EXPERIMENTS CAN BE DERIVED FROM THIS TABLE.

<table>
<thead>
<tr>
<th>AVIRIS Indian Pines</th>
<th>AVIRIS Kennedy Space Center</th>
<th>AVIRIS Salinas Valley</th>
<th>ROSIS Pavia University</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alfalfa (54)</td>
<td>Scrub (761)</td>
<td>Broccoli Green Weeds 1 (1893)</td>
<td>Asphalt (6631)</td>
</tr>
<tr>
<td>Corn-Notill (1434)</td>
<td>Willow (243)</td>
<td>Broccoli Green Weeds 2 (3704)</td>
<td>Meadows (18649)</td>
</tr>
<tr>
<td>Corn-Min (834)</td>
<td>Hammock (256)</td>
<td>Fallow (1960)</td>
<td>Gravel (2099)</td>
</tr>
<tr>
<td>Corn (234)</td>
<td>Oak (252)</td>
<td>Fallow Rough Plow (1228)</td>
<td>Trees (3064)</td>
</tr>
<tr>
<td>Grass-Pasture (497)</td>
<td>Slash Pine (161)</td>
<td>Fallow Smooth (2560)</td>
<td>Metal Sheets (1345)</td>
</tr>
<tr>
<td>Grass-Trees (747)</td>
<td>Oak/Broadleaf (229)</td>
<td>Stubble (3841)</td>
<td>Bare Soil (5029)</td>
</tr>
<tr>
<td>Grass-Pasture-Mowed (26)</td>
<td>Hardwood swamp (105)</td>
<td>Celery (3543)</td>
<td>Bitumen (1330)</td>
</tr>
<tr>
<td>Hay-Windrowed (489)</td>
<td>Graminoid marsh (311)</td>
<td>Grapes Untrained (11287)</td>
<td>Self-Blocking Bricks (3682)</td>
</tr>
<tr>
<td>Oats (20)</td>
<td>Spartina marsh (320)</td>
<td>Soil V irrigant (6128)</td>
<td>Shadow (947)</td>
</tr>
<tr>
<td>Soybeans-Notill (968)</td>
<td>Cattail Marsh (404)</td>
<td>Corn Senesce Green Weeds (3154)</td>
<td></td>
</tr>
<tr>
<td>Soybeans-Min (2468)</td>
<td>Salt marsh (186)</td>
<td>Lettuce Romaine 4 weeks (984)</td>
<td></td>
</tr>
<tr>
<td>Soybeans-Clean (614)</td>
<td>Mud flats (134)</td>
<td>Lettuce Romaine 5 weeks (1850)</td>
<td></td>
</tr>
<tr>
<td>Wheat (212)</td>
<td></td>
<td>Lettuce Romaine 6 weeks (818)</td>
<td></td>
</tr>
<tr>
<td>Woods (1294)</td>
<td></td>
<td>Lettuce Romaine 7 weeks (1003)</td>
<td></td>
</tr>
<tr>
<td>Bldg-Grass-Tree-Drives (380)</td>
<td></td>
<td>V irrigant Untrained (7055)</td>
<td></td>
</tr>
<tr>
<td>Stone-Steel-Towers (95)</td>
<td></td>
<td>V irrigant Vertical Trellis (1622)</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 3. (a) False color composition of the AVIRIS Indian Pines scene. (b) Ground truth-map containing 16 mutually exclusive land-cover classes (right).

B. AVIRIS Salinas Valley

The second AVIRIS data set used in experiments was collected over the Valley of Salinas in Southern California. The full scene consists of 512 lines by 217 samples with 186 spectral bands (after removal of water absorption and noisy bands) from 0.4 to 2.5 μm, nominal spectral resolution of 10 nm, and 16-bit radiometric resolution. It was taken at low altitude with a pixel size of 3.7 meters (high spatial resolution). The data include vegetables, bare soils and vineyard fields. Fig. 4(a) shows a false color composition of the scene and Fig. 4(b) shows the available ground-truth regions for this scene, which cover about two thirds of the entire Salinas scene. Finally, Fig. 4(c) shows some pictures of selected land-cover classes taken on the imaged site at the same time as the data was being collected by the sensor. Of particular interest are the relevant differences in the romaine lettuce classes resulting from different soil cover proportions.

C. AVIRIS Kennedy Space Center

The third data set used in experiments was collected by the AVIRIS sensor over the Kennedy Space Center,² Florida, on March 1996. The portion of this scene used in our experiments has dimensions of 292 × 383 pixels. After removing water absorption and low SNR bands, 176 bands were used for the analysis. The spatial resolution is 20 meters by pixel. 12 ground-truth classes where available, where the number of pixels in the smallest class is 105 while the number of pixels in the largest class is 761.

D. ROSIS Pavia

The fourth data set used in experiments was collected by the ROSIS optical sensor over the urban area of the University of Pavia, Italy. The flight was operated by the Deutschen Zentrum für Luftund Raumfahrt (DLR, the German Aerospace Agency) in the framework of the HySens project, managed and sponsored

²Available online: http://www.csr.utexas.edu/hyperspectral/data/KSC/.
IV. EXPERIMENTAL RESULTS

In this section we conduct a quantitative and comparative analysis of different feature extraction techniques for hyperspectral image classification, including unmixing-based and more traditional (supervised and unsupervised) approaches. The main goal is to use spectral unmixing and classification as complementary techniques, since the latter are more suitable for the classification of pixels dominated by a single land cover class, while the former are devoted to the characterization of mixed pixels. Because hyperspectral images often contain areas with both pure and mixed pixels, the combination of these two analysis techniques provides a synergistic data processing approach that has been explored in previous contributions [15], [27]–[30]. Before describing the results obtained in experimental validation, we first describe the feature extraction techniques that will be used in our comparison in Section IV-A. Then, Section IV-B describes the adopted supervised classification system and the experimental setup. Finally, Section IV-C discusses the obtained results in comparative fashion.

A. Feature Extraction Techniques Used in the Comparison

In our classification system, relevant features are first extracted from the original image. Several types of input features are used in the classification process, including spectral reflectance values and derived indices. These features are extracted from the input hyperspectral image and are used as input to the classification algorithms. The feature extraction process consists of two main steps: preprocessing and feature selection.

1. Preprocessing

The preprocessing step involves several operations such as radiometric correction, atmospheric correction, and normalization. These operations are performed to enhance the quality of the input data and to remove artifacts that may affect the accuracy of the classification results. For example, atmospheric correction is performed to remove the effects of atmospheric scattering and absorption on the observed spectral signatures.

2. Feature Selection

After preprocessing, the next step is to select the most relevant features for the classification task. This step is crucial because it helps to reduce the dimensionality of the data and to improve the performance of the classification algorithm. Feature selection can be performed using either supervised or unsupervised methods. In the supervised approach, the features are selected based on their discriminative power, while in the unsupervised approach, the features are selected based on their mutual information.

The selected features are then used as input to the classification algorithm. The algorithm is trained on the selected features and is used to classify the unknown pixels in the image. The classification results are then compared with the ground truth data to evaluate the performance of the algorithm.

By using a combination of feature extraction techniques, we can improve the accuracy of the classification results and obtain more reliable and accurate information about the land cover classes in the study area.
have been considered in the classification experiments conducted in this work. In the following, we provide an overview of the techniques used to extract features from the original hyperspectral data. A detailed mathematical description of these techniques is out of the scope of this work, since most of them are algorithms well known in the remote sensing literature, so only a short description of the conceptual basics for each method is given here. The techniques are divided into unsupervised approaches, if the algorithm is applied on the whole data cube, or supervised techniques, if the information associated with the training set of the data is somehow exploited during the feature extraction step.

1) Unsupervised Feature Extraction Techniques: We consider five unsupervised feature extraction techniques in this work. Three of them are classic algorithms available in the literature (PCA, MNF and ICA), and the two remaining ones are based on the exploitation of sub-pixel information through spectral unmixing, including the best unsupervised method in [15] and a newly proposed technique CMTMFMₜₚₚ in this work. A brief summary of the considered unsupervised techniques follows:

- **Principal component analysis** (PCA) is an orthogonal linear transformation which projects the data into new coordinate system, such that the greatest amount of variance of the original data is contained in the first principal components [11]. The resulting components are uncorrelated.

- **Minimum noise fraction** (MNF) differs from PCA in the fact that MNF ranks the obtained components according to their signal-to-noise ratio [9].

- **Independent component analysis** (ICA) tries to find components as statistically independent as possible, minimizing all the dependencies in the order up to fourth [10]. There are several strategies that can be adopted to define independence (e.g., minimization of mutual information, maximization of non-Gaussianity, etc.) In this work, among several possible implementations, we have chosen JADE [31] which provides a good tradeoff between performance and computational complexity when used for dimensionality reduction of hyperspectral images.

- **Unsupervised Mixture-Tuned Matched Filtering** (MTMFₜₚₚ), which first performs an MNF-based dimensionality reduction and then applies the MTMF method in order to estimate fractional abundances of spectral endmembers extracted from the original data using the orthogonal subspace projection (OSP) algorithm [32]. In [15] it is shown that MTMF outperforms other techniques for abundance estimation such as unconstrained and fully constrained linear spectral unmixing (FCLSU) [23] since it can provide meaningful abundance maps by means of partial unmixing in case not all endmembers are available a priori.

- **Unsupervised Clustering followed by Mixture-Tuned Matched Filtering** (CMTMFMₜₚₚ) developed in this work and intended to solve the problems highlighted by endmember extraction algorithms which are sensitive to outliers and pixels with extreme values of reflectance. By using an unsupervised clustering method such as the \( k \)-means to extract features, the endmembers extracted are expected to be more spatially significant.

- **Unsupervised Fuzzy Clustering** \( \text{FC}_{\text{unsup}} \) is an extension of the \( k \)-means clustering method [33] which provides soft clusters, where a particular pixel has a degree of membership in each cluster. This strategy is faster than the two previous strategies as it does not include a spectral unmixing step.

2) Supervised Feature Extraction Techniques: We consider several supervised feature extraction techniques in this work. The first techniques considered were discriminant analysis for feature extraction (DAFE) and decision boundary feature extraction (DBFE) [4]. However, DBFE could not be applied in the case of very limited training sets since it requires a number of samples (for each class) bigger than the number of dimensions of the original data set in order to estimate the statistics used to project the data. As it will be shown in the next sections, these requirement was not satisfied for most of the experiments carried out in this work. In turn, the results provided by DAFE were poor compared to the other methods for a low number of training samples, hence we did not include them in our comparison. As a result, the supervised methods adopted in our comparison were NWFE and three sub-pixel techniques based on estimating fractional abundances. Two of them were already presented in [15], and the third one is the CMTMFMₜₚₚ technique developed in this work. Although a number of supervised feature extraction techniques has been available in the literature [4], according to our experiments the advantages provided by supervised techniques is not always evident, especially in the case of limited training sets [34]. A brief summary of the considered supervised techniques follows:

- **Non-parametric weighted feature extraction** (NWFE) focuses on selecting samples near the eventual decision boundaries that best separate the classes. The main ideas of the NWFE are: 1) assigning different weights to every training sample in order to compute local means, and 2) defining non-parametric between-class and within-class scatter matrices to perform feature extraction [4].

- **Supervised Mixture-Tuned Matched Filtering** (M1MFₜₚₚ) is equivalent to M1MFₜₚₚ but assuming that the pure spectral components are searched by the OSP endmember extraction algorithm in the training set instead of in the entire hyperspectral image. Our assumption is that training samples may better represent the available land cover classes in the subsequent classification process [15].

- **Averaged Mixture-Tuned Matched Filtering** (M1MFₜₚₚ) is equivalent to M1MFₜₚₚ but assuming that the representative spectral signatures are obtained as the average of the signatures belonging to each class in the training set (here, the number of components to be retained by the MNF applied prior to the MTMF is varied in a given range). In this case, the OSP algorithm is not used to extract the spectral signatures, which are obtained in supervised fashion from the available training samples [15].

- **Supervised Clustering followed by Mixture-Tuned Matched Filtering** (CMTMFMₜₚₚ) developed in this work and acting as the supervised counterpart of
CMTMF \textsubscript{unsup}). It mainly differs with regards to that technique in the fact that the clustering process is performed in the training samples, and not in the full hyperspectral image.

### B. Supervised Classification System and Experimental Setup

In our supervised classification system, different types of input features are extracted from the original hyperspectral image prior to classification. In addition to the unsupervised and supervised feature extraction techniques described in the previous subsection, we also use the (full) original spectral information available in the hyperspectral data as input to the proposed classification system. In the latter case, the dimensionality of the input features used for classification equals \( n \), the number of spectral bands in the original data set. When using feature extraction techniques, the number of features was varied empirically in our experiments and only the best results are reported. In all cases, a supervised classification process was performed using the SVM classifier with Gaussian kernel (observed to perform better than other tested kernels, such as polynomial or linear). Kernel parameters were optimized by a grid search procedure, and the optimal parameters were selected using 10-fold cross-validation (selected after testing different configurations). The LIBSVM library\(^3\) was in our experiments. In order to evaluate the ability of the tested methods to perform under training sets with different number of samples, we adopted the following training-test configurations:

- In our experiments with the AVIRIS Indian Pines data set in Fig. 3(a), we randomly selected 5% and 15% of the pixels in each ground-truth class in Table I and used them to build the training set. The remaining pixels were used as test pixels.
- In our experiments with the AVIRIS Salinas data set in Fig. 4(a), in which the size of the smaller classes is bigger when compared to those in the AVIRIS Indian Pines data set, we decided to reduce the training sets even more and selected only 2% and 5% of the available ground-truth pixels in Table I for training purposes.
- In our experiments with the AVIRIS Kennedy Space Center data set, we decided to reduce the training sets even more and selected only 1% and 5% of the available ground-truth pixels in Table I for training purposes.
- Finally, in our experiments with the ROSIS Pavia data set in Fig. 5(a), we used the training set in Fig. 5(c) and also a different training set made up of only 50 pixels for each class in Table I for comparative purposes.

Based on the aforementioned training sets, the overall (OA) and average (AA) classification accuracies were computed over the remaining test samples for each data set. This experiment was repeated ten times to guarantee statistical consistency, and the average results after ten runs are provided. An assessment of the obtained results is reported in the following subsection.

### C. Analysis and Discussion of Results

Table II shows the OA and AA (in percentage) obtained by the considered classification system for different hyperspectral scenes using the original spectral information as input feature, and also the features provided by the unsupervised and supervised feature extraction techniques described in Section IV-A. It is important to emphasize that, in the tables, we only report the best case (meaning the one with highest OA) for each considered feature extraction technique, after testing numbers of extracted features ranging from 5 to 50. In all cases, this range was sufficient to observe a decline in classification OA after a certain number of features, so the number given in the parentheses in the tables correspond to the optimal number of features for each considered feature extraction technique (in the case of the original spectral information, the number in the parentheses corresponds to the number of bands of the original hyperspectral image). Finally, in order to outline the best feature extraction

\(^3\)http://www.csie.ntu.edu.tw/cjlin/libsvm/.
From Table II, several conclusions can be drawn. First and foremost, we can observe that the use of supervised techniques for feature extraction is not always beneficial to improve the OA and AA, especially in case of limited training sets and statistical feature extraction approaches. For example, NWFE exhibits better results when compared to traditional unsupervised techniques such as PCA or ICA. However, DAFE (not included in the tables) exhibited quite poor results. The low performances obtained by DAFE should be therefore attributed to the very small size of the training set and to the fact that the land cover classes can be spectrally very close (as in the case of the AVIRIS Indian Pines scene) thus making it very difficult to separate them by using spectral means and covariance matrices. Moreover, the importance of integrating the additional information provided by the training samples is strictly connected with the nature of the considered approach. This can be noticed when comparing the MTMF versus the CMTMF chains. In the former case, the best results are generally provided by the supervised approach since the supervised strategy for extracting spectral endmembers using the OSP approach benefits from the reduction of outliers and pixels with extreme values of reflectance, which affect negatively this endmember extraction algorithm. In the latter case, the best results are generally provided by the unsupervised approach due to the fact that, when trying to identify clusters in a very small training set, several problems appear, such as the bad conditioning of matrices when computing the inverse (in the -means clustering step) or the eventual selection of very similar clusters, leading to redundant information in class prototyping which ultimately affects the subsequent partial unmixing step and the obtained classification performances. In addition to the aforementioned observations, we emphasize that the supervised version derives the endmembers (via clustering) from a limited training set, while the unsupervised version derives the endmembers from
the whole hyperspectral image. The former approach has the advantage of computational complexity, as the search for endmembers is only conducted in the small training set, but this comes at the expense of reduced modelling accuracy as expected. Although in previous work we developed MTMF$_{\text{avg}}$ in the hope of addressing these problems, our experimental results in this work indicate that CMTMF techniques in general and CMTMF$_{\text{unsup}}$ in particular (an unsupervised approach as opposed to MTMF$_{\text{avg}}$) performs a better job in characterizing the sub-pixel information prior to classification of hyperspectral data. Finally, it is also worth noting the good performance achieved in all experiments by MNF, another unsupervised feature extraction strategy. Figs. 6–8 show the results obtained in some of the experiments.

An arising question at this point is whether there is any advantage of using unmixing chains versus the MNF transform. Since both feature extraction methods are unsupervised, with similar computational complexity and leading to similar classification results, it is not clear from the context if there exists any advantage of using an unmixing-based technique over a well-known, statistical method such as the MNF. In order to address this issue, Fig. 9 shows the first nine components extracted by the MNF from the ROSIS Pavia University image. These components are ordered in terms of signal-to-noise ratio, with the first component providing the maximum amount of information. Here, noise can be clearly appreciated in the last three components. In turn, Fig. 10 shows the components extracted for the same image by the CMTMF$_{\text{unsup}}$ technique. The components are arranged in no specific order, as spectral unmixing assigns the same priority to each endmember when deriving the associated abundance map. As shown by Fig. 10, the components provided by the unmixing-based technique can be interpreted in a physical manner (as the abundances of each spectral constituent in the scene) and most importantly these components can be related to the ground-truth classes in Fig. 5(a). This suggests that unmixing-based chains can provide an alternative strategy to classic feature extraction chains such as the MNF with three main differences:

1) Unmixing-based feature extraction techniques incorporate information about mixed pixels, which are the dominant type of pixel in hyperspectral images. Quite opposite, standard feature extraction techniques such as the MNF do not incorporate the pure/mixed nature of the pixels in hyperspectral data, disregarding a source of information that could be useful for the final classification.

2) The components provided by unmixing-based feature extraction techniques can be interpreted as the abundance of spectral constituents in the scene, while the components provided by other classic feature extraction techniques such as the MNF do not necessarily have any physical meaning.

3) Unmixing-based feature extraction techniques do not penalize classes which are not relevant in terms of variance or signal-to-noise ratio, while some classic feature extraction techniques such as the MNF relegate variations of less significant size to low-order components. If such low-order components are not preserved, small classes may be affected.

An additional aspect resulting from our experiments is that unmixing-based chains allow for a natural integration of the spatial information available in the original hyperspectral image (through the clustering strategy for endmember extraction designed in this work). Although the aforementioned aspects may offer important advantages in hyperspectral data classification, the true fact is that our comparative assessment (conducted in terms of OA and AA using four representative hyperspectral images) only indicates a moderate improvement (or comparable performance) of the best unmixing-guided feature extraction method (CMTMF$_{\text{unsup}}$) with regards to the best statistical feature extraction method (MNF) reported in our experiments. This leads us to believe that further improvements to the integration of the information provided by spectral unmixing into the classification process are possible. With this in mind, we anticipate significant advances in the integration of spectral unmixing and classification of hyperspectral data in future developments.

V. CONCLUSIONS AND FUTURE LINES

In this paper, we have investigated the advantages that can be gained by including information about spectral mixing at sub-pixel levels in the feature extraction stage that is usually conducted prior to hyperspectral image classification. For this purpose, we have developed a new unmixing-based feature extraction technique that combines the spatial and the spectral information through a combination of unsupervised clustering and partial spectral unmixing. We have compared our newly developed technique (which can be applied in both unsupervised and supervised fashion) with other classic and unmixing-based techniques for feature extraction. Our detailed quantitative and comparative assessment has been conducted using four representative hyperspectral images collected by two different instruments (AVIRIS and ROSIS) over a variety of test sites and in the framework of supervised classification scenarios dominated by the limited availability of training samples. Our experimental results indicate that the unsupervised version of our newly developed technique provides components which are physically meaningful and significant from a spatial point of view, resulting in good classification accuracies (without penalizing very small classes) when compared to the other feature extraction techniques tested in this work. In turn, since our analysis scenarios are dominated by very limited training sets, we have experimentally observed that, in this context, the use of supervised feature extraction techniques can lead to lower classification accuracies as the information considered for projecting the data into a lower-dimensional space is not representative of the thematic classes of the image.

Future developments of this work will include an investigation of additional techniques for feature extraction from a spectral unmixing point of view, in order to fully substantiate the advantages that can be gained at the feature extraction stage by including additional information about mixed pixels (which are predominant in hyperspectral images) prior to classification purposes. Another research line deserving future attention is the determination of automatic procedures to determine the optimal number of features to be extracted from each tested method.
While methods for estimating the intrinsic dimensionality of hyperspectral images exist, the determination of the number of features suitable for classification purposes depends on each particular method and, in the case of supervised feature extraction methods, on the available training. Although in this work we have investigated performance in a suitable range of extracted features, the automatic determination of the optimal number of features for each method should be investigated in future work for practical reasons. Finally, future work should also consider nonlinear feature extraction methods such as kernel PCA [37] in addition to the linear feature extraction methods considered in this work.

ACKNOWLEDGMENT

The authors would also like to thank D. Landgrebe, M. Crawford, and L. F. Johnson for sharing the hyperspectral data sets used in this work.
REFERENCES


Inmaculada Dópido received the B.S. and M.S. degrees in telecommunications from the University of Extremadura, Caceres, Spain, where she is currently working towards the Ph.D. degree. She is a member of the Hyperspectral Computing Laboratory (HypeComp) coordinated by Prof. Antonio Plaza. Her research interests include remotely sensed hyperspectral imaging, pattern recognition and signal and image processing, with particular emphasis on the development of new techniques for unsupervised and supervised classification and spectral mixture analysis of hyperspectral data.

Ms. Dópido has been a manuscript reviewer for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.

Alberto Villa (S’09–M’11) received the B.S. and M.S. degrees in electronic engineering from the University of Pavia, Pavia, Italy, in 2005 and 2008, respectively. In 2011, he received the Ph.D. degree (a joint degree) from the Grenoble Institute of Technology (Grenoble INP), Grenoble, France, and the University of Iceland, Reykjavik, Iceland. He was a visiting researcher at the Hyperspectral Computing Laboratory (HypeComp), University of Extremadura, Spain, from September 2010 to February 2011. Since July 2011, he has been working as a research engineer for Aresys srl, a spin-off company of Politecnico di Milano.
Fig. 7. Classification results for the AVIRIS Salinas Valley scene (obtained using an SVM classifier with Gaussian kernel, trained with 2% of the available samples). (a) Ground Truth; (b) PCA; (c) ICA; (d) MNF; (e) MTMF$_{d_1}$; (f) CMTMF$_{d_1}$; (g) NWFE; (h) MTMF$_{d_2}$; (i) MTMF$_{avg}$; (j) CMTMF$_{avg}$.

dealing with SAR imaging and ground-based radar. His research interests are in the areas of SAR antenna model, spectral unmixing, machine learning, hyperspectral imaging, signal and image processing.

Dr. Villa is a reviewer for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING.
Antonio Plaza (M’05–SM’07) received the M.S. and Ph.D. degrees in computer engineering from the University of Extremadura, Caceres, Spain. He was a Visiting Researcher with the Remote Sensing Signal and Image Processing Laboratory, University of Maryland Baltimore County, Baltimore, with the Applied Information Sciences Branch, Goddard Space Flight Center, Greenbelt, MD, and with the AVIRIS Data Facility, Jet Propulsion Laboratory, Pasadena, CA. He is currently an Associate Professor with the Department of Technology of Computers and Communications, University of Extremadura, Caceres, Spain, where he is the Head of the Hyperspectral Computing Laboratory (HyperComp). He was the Coordinator of the Hyperspectral Imaging Network (Hyper-I-Net), a European project designed to build an interdisciplinary research community focused on hyperspectral imaging activities. He has been a Proposal Reviewer with the European Commission, the European Space Agency, and the Spanish Government. He is the author or coauthor of around 300 publications on remotely sensed hyperspectral imaging, including more than 60 Journal Citation Report papers, 20 book chapters, and over 200 conference proceeding papers. His research interests include remotely sensed hyperspectral imaging, pattern recognition, signal and image processing, and efficient implementation of large-scale scientific problems on parallel and distributed computer architectures.

Dr. Plaza has coedited a book on high-performance computing in remote sensing and guest edited seven special issues on remotely sensed hyperspectral imaging for different journals, including the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (for which he serves as Associate Editor on hyperspectral image analysis and signal processing since 2007), the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, the International Journal of High Performance Computing Applications, and the Journal of Real-Time Image Processing. He is also serving as an Associate Editor for the IEEE GEOSCIENCE AND REMOTE SENSING NEWSLETTER. He has served as a reviewer for more than 280 manuscripts submitted to more than 50 different journals, including more than 140 manuscripts reviewed for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He has served as a Chair for the IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing in 2011. He has also been serving as a Chair for the SPIE Conference on Satellite Data Compression, Communications, and Processing since 2009, and for the SPIE Remote Sensing Europe Conference on High Performance Computing in Remote Sensing since 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. He is currently serving as Director of Education activities and member of the Administrative Committee of the IEEE Geoscience and Remote Sensing Society.

Paolo Gamba (M’93–SM’00) is currently an Associate Professor of telecommunications at the University of Pavia, Italy. Since January 2009 he serves as Editor-in-Chief of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. He also served as Technical Co-Chair of the 2010 IEEE Geoscience and Remote Sensing Symposium, Honolulu, Hawaii, July 2010. He has been the organizer and Technical Chair of the biennial GRSS/ISPRS Joint Workshops on “Remote Sensing and Data Fusion over Urban Areas” since 2001. The next conference in the series, called JURSE 2013, is going to be Sao Paulo in 2013. He has been Chair of Technical Committee 7 “Pattern Recognition in Remote Sensing” of the International Association for Pattern Recognition (IAPR) from October 2002 to October 2004 and Chair of the Data Fusion Committee of the IEEE Geoscience and Remote Sensing Society from October 2005 to May 2009.

Dr. Gamba has been the Guest Editor of special issues of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, ISPRS Journal of Photogrammetry and Remote Sensing, International Journal of Information Fusion and Pattern Recognition Letters on the topic of Urban Remote Sensing, Remote Sensing for Disaster Management, Pattern Recognition in Remote Sensing Applications. He has been invited to give keynote lectures and tutorials in many international conferences. He published more than 80 papers on international peer-review journals and presented more than 210 papers in workshops and conferences.
Fig. 8. Classification results for the ROSIS Pavia University scene (obtained using an SVM classifier with Gaussian kernel, trained with 50 pixels of each available ground-truth class). (a) Ground Truth; (b) PCA; (c) ICA; (d) MNF; (e) $\text{MTMF}_{\text{unmix}}$; (f) $\text{CMTMF}_{\text{unmix}}$; (g) NWFE; (h) $\text{MTMF}_{\text{avg}}$; (i) $\text{MTMF}_{\text{avg}}$; (j) $\text{CMTMF}_{\text{avg}}$. 
Fig. 9. Components extracted by the MNF from the ROSIS Pavia University scene (ordered from left to right in terms of amount of information).

Fig. 10. Components extracted by the CMTMF feature extraction technique from the ROSIS Pavia University scene (in no specific order).