Spectral–Spatial Hyperspectral Image Segmentation Using Subspace Multinomial Logistic Regression and Markov Random Fields

Jun Li, José M. Bioucas-Dias, *Member, IEEE*, and Antonio Plaza, *Senior Member, IEEE*

Abstract—This paper introduces a new supervised segmentation algorithm for remotely sensed hyperspectral image data which integrates the spectral and spatial information in a Bayesian framework. A multinomial logistic regression (MLR) algorithm is first used to learn the posterior probability distributions from the spectral information, using a subspace projection method to better characterize noise and highly mixed pixels. Then, contextual information is included using a multilevel logistic Markov–Gibbs Markov random field prior. Finally, a maximum *a posteriori* segmentation is efficiently computed by the α-Expansion min-cut-based integer optimization algorithm. The proposed segmentation approach is experimentally evaluated using both simulated and real hyperspectral data sets, exhibiting state-of-the-art performance when compared with recently introduced hyperspectral image classification methods. The integration of subspace projection methods with the MLR algorithm, combined with the use of spatial–contextual information, represents an innovative contribution in the literature. This approach is shown to provide accurate characterization of hyperspectral imagery in both the spectral and the spatial domain.

Index Terms—Hyperspectral image segmentation, Markov random field (MRF), multinomial logistic regression (MLR), subspace projection method.

I. INTRODUCTION

SUPERVISED classification (and segmentation) of high-dimensional data sets such as remotely sensed hyperspectral images is a difficult endeavor [1]. Obstacles, such as the Hughes phenomenon [2], appear as the data dimensionality increases. This is because the number of training samples used for the learning stage of the classifier is generally very limited compared with the number of available spectral bands. In order to circumvent this problem, several feature selection [3] and extraction [4] methods have been combined with machine learning techniques that are able to perform accurately in the presence of limited training sets, including support vector machines (SVMs) [5], [6] or multinomial logistic regression (MLR)-based classifiers [7], [8].

Due to sensor design considerations, the wealth of spectral information in hyperspectral data is often not complemented by extremely fine spatial resolution. This (and other phenomena, such as the presence of mixtures of components at different scales) leads to the problem of mixed pixels, which represent a challenge for accurate hyperspectral image classification [9]. In order to address this issue, subspace projection methods [10] have been shown to be a powerful class of statistical pattern classification algorithms [11]. These methods can handle the high dimensionality of an input data set by bringing it to the right subspace without losing the original information that allows for the separation of classes.

In this context, subspace projection methods can provide competitive advantages by separating classes which are very similar in spectral sense, thus addressing the limitations in the classification process due to the presence of highly mixed pixels. The idea of applying subspace projection methods to improve classification relies on the basic assumption that the samples within each class can approximately lie in a lower dimensional subspace. Thus, each class may be represented by a subspace spanned by a set of basis vectors, while the classification criterion for a new input sample would be the distance from the class subspace [12]–[14]. Recently, several subspace projection methods have been specifically designed for improving hyperspectral data characterization [3], [15]–[17], obtaining successful results. Another recent trend is to combine spectral and spatial–contextual information [8], [9], [18]–[22]. In some of these works, Markov random fields (MRFs) have obtained great success in characterizing spatial information in remotely sensed data sets [23], [24]. MRFs exploit the continuity, in probability sense, of neighboring labels. The basic assumption is that, in a hyperspectral image, it is very likely that two neighboring pixels will have the class same label.

In this paper, we propose a new Bayesian approach to hyperspectral image segmentation which combines spectral and spatial information. The algorithm implements the following two main steps: 1) learning, where the posterior probability distributions are modeled by an MLR combined with a subspace projection method, and 2) segmentation, which infers an image of class labels from a posterior distribution built on the learned subspace classifier and on a multilevel logistic (MLL) prior on the image of labels. The final output of the algorithm is based

on a maximum *a posteriori* (MAP) segmentation process which is computed via an efficient min-cut-based integer optimization technique. The main novelty of our proposed work is the integration of a subspace projection method with the MLR which is further combined with spatial–contextual information, which will be shown to provide a good characterization of content of hyperspectral imagery in both the spectral and the spatial domain. The proposed Bayesian method exhibits good discriminatory capability when dealing with ill-posed problems, i.e., limited training samples versus high dimensionality of the input data. In addition to this, we emphasize that the proposed approach provides class posterior probabilities which are crucial to the complete posterior probabilities, such that the final MAP segmentation can benefit from the inclusion of both the spectral and the spatial information available in the original hyperspectral data. As will be shown by our experimental results, the accuracies achieved by our approach are competitive or superior to those provided by many other state-of-the-art supervised classifiers for hyperspectral image analysis. Furthermore, an important innovative contribution of this work with regard to our previous work in [8] and [22] is that the proposed method applies a subspace projection method instead of considering the full spectral information as an input to the MLR model, thus being able to circumvent some limitations in the techniques described in [8] and [22] due to the high dimensionality of the input data and the presence of highly mixed pixels, which are now tackled simultaneously by our newly proposed approach.

The remainder of this paper is organized as follows. Section II formulates the problem. Section III describes the proposed approach. Section IV reports segmentation results based on simulated and real hyperspectral data sets in comparison with other state-of-the-art supervised classifiers. Finally, Section V concludes with some remarks and hints at plausible future research lines.

II. PROBLEM FORMULATION

Before describing our proposed approach, let us first define some of the notations that will be used throughout this paper:

$\mathcal{S} \equiv \{1, \ldots, n\}$	Set of integers indexing the n pixels of an image.
$\mathcal{K} \equiv \{1, \ldots, K\}$	Set of K classes.
$\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathbb{R}^{d \times n}$	Image in which the pixels are d-dimensional vectors.
$\mathbf{y} = (y_1, \ldots, y_n) \in \mathcal{L}^n$	Image of labels.
$\mathcal{D}_{l^{(k)}}^{(k)} \equiv \{(y_1, \mathbf{x}_1), \ldots, (y_{l^{(k)}}, \mathbf{x}_{l^{(k)}})\}$	Set of labeled samples for class k with size $l^{(k)}$.
$\mathbf{x}_{l^{(k)}}^{(k)} \equiv \{\mathbf{x}_1, \ldots, \mathbf{x}_{l^{(k)}}\}$	Set of feature vectors in $\mathcal{D}_{l^{(k)}}^{(k)}$.
$\mathcal{D}_l \equiv \{\mathcal{D}_{l^{(1)}}^{(1)}, \ldots, \mathcal{D}_{l^{(K)}}^{(K)}\}$	Set of labeled samples with size $l = \sum_{k=1}^{K} l^{(k)}$.

With the aforementioned definitions in place, the goal of classification is to assign a label $y_i \in \mathcal{K}$ to each pixel vector $\mathbf{x}_i$, with $i \in \mathcal{S}$. This process results in an image of class labels $\mathbf{y}$, and we will call this assignment *labeling*. In turn, the goal of segmentation is to partition the set $\mathcal{S}$ such that the pixels in each subset $\mathcal{S}_k$, with $\mathcal{S} = \cup_k \mathcal{S}_k$, share some common property, e.g., they represent the same type of land cover. Notice that, given a labeling $\mathbf{y}$, the collection $\mathcal{S}_k = \{i \in \mathcal{S} | y_i = k\}$ for $k \in \mathcal{K}$ is a partition of $\mathcal{S}$. On the other hand, given the segmentation $\mathcal{S}_k$ for $k \in \mathcal{K}$, the image $\{y_i | y_i = k, \text{ if } i \in \mathcal{S}_k, i \in \mathcal{S}\}$ is a labeling. As a result, there is a one-to-one relationship between labelings and segmentations. Without loss of generality, in this paper, we use the term *classification* when the spatial information in the original scene is not used in the labeling process. Similarly, we use the term *segmentation* when the spatial information in the original scene is used to produce such labeling.

In a Bayesian framework, the labeling process[1] is usually conducted by maximizing the posterior distribution[1] as follows:

$$p(\mathbf{y}|\mathbf{x}) \propto p(\mathbf{x}|\mathbf{y})p(\mathbf{y}) \tag{1}$$

where $p(\mathbf{x}|\mathbf{y})$ is the likelihood function (i.e., the probability of the feature image given the labels) and $p(\mathbf{y})$ is the prior over the image of labels. Assuming conditional independence of the features given the labels, i.e., $p(\mathbf{x}|\mathbf{y}) = \prod_{i=1}^{i=n} p(\mathbf{x}_i|y_i)$, then the posterior may be written as a function of $\mathbf{y}$ as follows:

$$\begin{aligned} p(\mathbf{y}|\mathbf{x}) &= \frac{1}{p(\mathbf{x})} p(\mathbf{x}|\mathbf{y}) p(\mathbf{y}) \\ &= \frac{1}{p(\mathbf{x})} \prod_{i=1}^{i=n} p(\mathbf{x}_i|y_i) p(\mathbf{y}) \\ &= \alpha(\mathbf{x}) \prod_{i=1}^{i=n} \frac{p(y_i|\mathbf{x}_i)}{p(y_i)} p(\mathbf{y}) \end{aligned} \tag{2}$$

where $\alpha(\mathbf{x}) \equiv \prod_{i=1}^{i=n} p(\mathbf{x}_i)/p(\mathbf{x})$ is a factor not depending on $\mathbf{y}$. In the proposed approach, we assume the classes are equally likely, i.e., $p(y_i = k) = 1/K$ for any $k \in \mathcal{K}$. However, any other distribution can be accommodated, as long as the marginal of $p(\mathbf{y})$ is compatible with the assumed distribution. Therefore, the MAP segmentation is given by

$$\widehat{\mathbf{y}} = \arg \max_{y \in \mathcal{K}^n} \left\{ \sum_{i=1}^{n} (\log p(y_i|\mathbf{x}_i) + \log p(\mathbf{y}) \right\}. \tag{3}$$

Following the Bayesian framework described earlier, we have developed a new algorithm which naturally integrates the spectral and the spatial information contained in the original hyperspectral image data. In our proposed algorithm, the spectral information is represented by class densities $p(y_i|\mathbf{x}_i)$, which are learned by a subspace projection-based MLR algorithm. On the other hand, the spatial prior $p(\mathbf{y})$ is given by an MRF-based MLL prior which encourages neighboring pixels to have the same label. The MAP segmentation $\widehat{\mathbf{y}}$ is computed via the

---

[1]To keep the notation simple, we use $p(\cdot)$ to denote both continuous densities and discrete distributions of random variables. The meaning should be clear from the context.

α-Expansion algorithm [25], a min-cut-based tool to efficiently solve integer optimization problems. Additional details are given in the following section.

III. PROPOSED APPROACH

This section is organized as follows. First, we propose a new spectral classifier which uses an MLR-based model combined with a subspace projection method. Second, we incorporate spatial information in the spectral classifier by resorting to an MRF-based isotropic MLL prior. Then, we describe an efficient min-cut optimization technique (α-Expansion algorithm) intended for computing a MAP segmentation. Finally, we integrate the individual modules described previously into a supervised Bayesian spatial–spectral segmentation framework.

A. Subspace Projection-Based MLR Classifier: MLRsub

One of the main problems in hyperspectral image classification is the presence of mixed pixels, which can be an important source of inaccuracies in the data interpretation process. Two models have been widely used in the literature for characterizing mixed pixels: linear and nonlinear [26]. The linear mixture model assumes that the spectral response of a mixed pixel can be explained as a linear combination of a set of pure spectral signatures (also called *endmembers*) weighted by their corresponding abundance fractions [27]. This model assumes minimal secondary reflections and/or multiple scattering effects in the data, as opposed to nonlinear unmixing which assumes that the endmembers form an intimate mixture inside the respective pixel. In the latter case, incident radiation interacts with more than one component and is affected by multiple scattering effects [28]. Nonlinear unmixing generally requires prior knowledge about the object geometry and physical properties of the observed objects [29], which may be very difficult to obtain in practice. In this paper, we focus exclusively on the linear mixture model, due to its computational tractability and flexibility in different applications.

Under the linear mixture model assumption, for any $i \in \mathcal{S}$, we have

$$\mathbf{x}_i = \mathbf{m}\boldsymbol{\gamma}_i + \mathbf{n}_i \qquad (4)$$

where $\mathbf{m} \equiv [\mathbf{m}^{(1)}, \ldots, \mathbf{m}^{(K)}]$ denotes a mixing matrix composed by the spectral endmembers, $\mathbf{n}_i$ denotes the noise, and $\boldsymbol{\gamma}_i = [\gamma_i^{(1)}, \ldots, \gamma_i^{(K)}]^{\mathrm{T}}$ denotes the fractional abundances of the endmembers in the mixed pixel $\mathbf{x}_i$. Since the distribution $p(\boldsymbol{\gamma}_i)$ is unknown, we cannot compute $p(\mathbf{x}_i|y_i = k)$ using a generative model. It happens, however, that the linear term $\mathbf{m}\boldsymbol{\gamma}_i$ in (4) lives in class-dependent subspaces. This is a consequence of the linearity of this term and of the fact that the set of materials corresponding to any two different classes are very likely to be different. An important advantage of the proposed approach is the inclusion of a subspace projection method which takes advantage of the fact that, in general, hyperspectral data live in a much lower dimensional subspace compared with the feature dimensionality. Therefore, if the size of the training set for
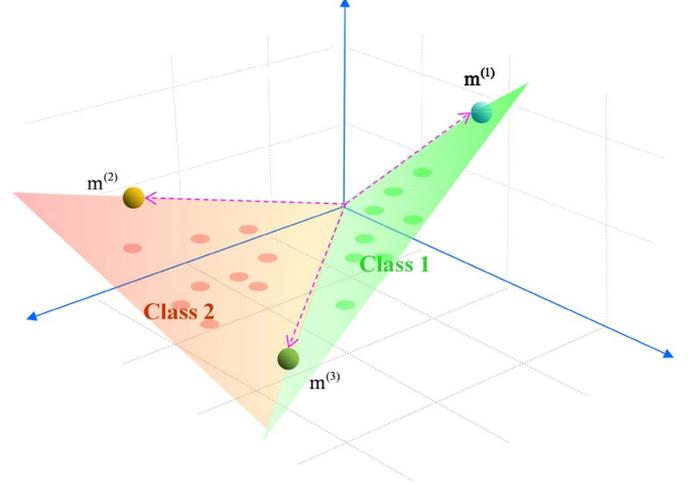
Fig. 1. Illustration of the advantages that can be gained by using subspace projection under the linear mixture model assumption, where $\{\mathbf{m}^{(1)}, \mathbf{m}^{(2)}, \mathbf{m}^{(3)}\}$ denote the spectral endmembers, to reduce the impact of mixed pixels in the classification process.

a certain class is very small, our proposed approach can still provide good performance in such ill-posed problems, provided that the method is able to obtain a good estimate of the subspace in which the considered class lives. Fig. 1 shows the advantages that can be gained by using subspace projection under the linear mixture model assumption. As shown by Fig. 1, hyperspectral data generally live in class-independent subspaces given by the spectral endmembers. Subsequently, the projection into such subspaces allows us to specifically avoid spectral confusion due to mixed pixels, thus reducing their impact in the subsequent classification process.

With this in mind, we may then write the observation mechanism for class k as

$$\mathbf{x}_i^{(k)} = \mathbf{U}^{(k)}\mathbf{z}_i^{(k)} + \mathbf{n}_i^{(k)} \qquad (5)$$

where $\mathbf{n}_i^{(k)}$ is the noise of class k, $\mathbf{U}^{(k)} = \{\mathbf{u}_1^{(k)}, \ldots, \mathbf{u}_{r^{(k)}}^{(k)}\}$ is a set of $r^{(k)}$-dimensional orthonormal basis vectors for the subspace associated with class k, and $\mathbf{z}_i^{(k)}$ is, apart from the noise $\mathbf{n}_i^{(k)}$, the coordinates of $\mathbf{x}_i^{(k)}$ with respect to the basis $\mathbf{U}^{(k)}$.

We assume that the class-independent random vectors $\mathbf{n}_i^{(k)}$ and $\mathbf{z}_i^{(k)}$ are Gaussian distributed with zero mean and diagonal covariance matrices, i.e., $\mathbf{n}_i^{(k)} \sim \mathcal{N}(0, \sigma^{(k)2}\mathbf{I})$ and $\mathbf{z}_i^{(k)} \sim \mathcal{N}(0, \alpha^{(k)}\mathbf{I})$. We are aware that, statistically, these assumptions are very strong and that they rarely hold in real data. However, they allow us to preserve the subspace-based structure of our model and yield a robust discriminative model. Hence, we have decided to preserve these assumptions in our implementation. In addition to this, we also emphasize that, accordingly, normalization of the input data is not needed in order for the proposed model to perform properly. Based on the aforementioned assumptions, we have the following generative model:

$$p(\mathbf{x}_i|y_i = k) \sim \mathcal{N}\left(0, \alpha^{(k)}\mathbf{U}^{(k)}\mathbf{U}^{(k)\mathrm{T}} + \sigma^{(k)2}\mathbf{I}\right). \qquad (6)$$

Under the present setup, the generative model in (6) can be computed as follows:

$$
\begin{aligned}
p(\mathbf{x}_i | y_i = k) \\
\propto \exp\Bigg\{ &-\frac{1}{2}\mathbf{x}_i^{\mathrm{T}} \left(\alpha^{(k)}\mathbf{U}^{(k)}\mathbf{U}^{(k)\mathrm{T}} + \sigma^{(k)2}\mathbf{I} \right)^{-1}\mathbf{x}_i \Bigg\} \\
= \exp\Bigg\{ &-\frac{1}{2}\mathbf{x}_i^{\mathrm{T}} \left(\frac{\mathbf{I}}{\sigma^{(k)2}} - \frac{\mathbf{U}^{(k)}}{\sigma^{(k)2}} \right. \\
&\times \left. \left(\frac{\mathbf{I}}{\alpha^{(k)}} + \frac{\mathbf{U}^{(k)\mathrm{T}}\mathbf{U}^{(k)}}{\sigma^{(k)2}} \right)^{-1}\frac{\mathbf{U}^{(k)\mathrm{T}}}{\sigma^{(k)2}} \right)\mathbf{x}_i \Bigg\} \\
= \exp\Bigg\{ &-\frac{1}{2\sigma^{(k)2}}\mathbf{x}_i^{\mathrm{T}}\left(1 - \frac{\alpha^{(k)}}{\alpha^{(k)}+\sigma^{(k)2}}\mathbf{U}^{(k)}\mathbf{U}^{(k)\mathrm{T}} \right)\mathbf{x}_i \Bigg\} \\
= \exp\Bigg\{ &-\frac{1}{2}\frac{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_i}{\sigma^{(k)2}} + \frac{1}{2\sigma^{(k)2}}\frac{\alpha^{(k)}}{\alpha^{(k)}+\sigma^{(k)2}}\left\| \mathbf{x}_i^{\mathrm{T}}\mathbf{U}^{(k)} \right\|^2 \Bigg\}. \quad (7)
\end{aligned}
$$

Let $\omega_1^{(k)} \equiv -(1/2\sigma^{(k)2})$, $\omega_2^{(k)} \equiv (1/2\sigma^{(k)2})(\alpha^{(k)}/(\alpha^{(k)} + \sigma^{(k)2}))$, $\boldsymbol{\omega}^{(k)} \equiv [\omega_1^{(k)} \ \omega_2^{(k)}]^{\mathrm{T}}$, and $\boldsymbol{\omega} \equiv [\boldsymbol{\omega}^{(1)\mathrm{T}}, \dots, \boldsymbol{\omega}^{(K)\mathrm{T}}]^{\mathrm{T}}$. With these definitions in mind, we can compute the posterior class density $p(y_i|\mathbf{x}_i)$ as follows:

$$
\begin{aligned}
p(y_i = k | \mathbf{x}_i, \boldsymbol{\omega}) &= \frac{p(\mathbf{x}_i|y_i = k, \boldsymbol{\omega})p(y_i = k)}{\sum_{k=1}^K p(\mathbf{x}_i|y_i = k, \boldsymbol{\omega})p(y_i = k)} \\
&= \frac{\exp\left(\boldsymbol{\omega}^{(k)\mathrm{T}}\boldsymbol{\phi}^{(k)}(\mathbf{x}_i) \right)p(y_i = k)}{\sum_{k=1}^K \exp\left(\boldsymbol{\omega}^{(k)\mathrm{T}}\boldsymbol{\phi}^{(k)}(\mathbf{x}_i) \right)p(y_i = k)} \quad (8)
\end{aligned}
$$

where $\boldsymbol{\phi}^{(k)}(\mathbf{x}_i) = [\|\mathbf{x}_i\|^2, \|\mathbf{x}_i^{\mathrm{T}}\mathbf{U}^{(k)}\|^2]^{\mathrm{T}}$. Assuming equiprobable classes, i.e., $p(y_i = k) = 1/K$, the problem in (8) turns to

$$
p(y_i = k | \mathbf{x}_i, \boldsymbol{\omega}) = \frac{\exp\left(\boldsymbol{\omega}^{(k)}\boldsymbol{\phi}^{(k)}(\mathbf{x}_i) \right)}{\sum_{k=1}^K \exp\left(\boldsymbol{\omega}^{(k)}\boldsymbol{\phi}^{(k)}(\mathbf{x}_i) \right)} \quad (9)
$$

which is exactly an MLR [7].

1) Learning the Class-Independent Subspace: Let $\mathbf{R}^{(k)} = \langle \mathbf{x}_{l^{(k)}}^{(k)}\mathbf{x}_{l^{(k)}}^{(k)\mathrm{T}} \rangle$ denote the sample correlation matrix associated with class k determined from the training set. By computing the eigendecomposition of $\mathbf{R}^{(k)}$, we have

$$
\mathbf{R}^{(k)} = \mathbf{E}^{(k)}\boldsymbol{\Lambda}^{(k)}\mathbf{E}^{(k)\mathrm{T}} \quad (10)
$$

where $\mathbf{E}^{(k)} = \{\mathbf{e}_1^{(k)}, \dots, \mathbf{e}_d^{(k)}\}$ is the eigenvector matrix and $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1^{(k)}, \dots, \lambda_d^{(k)})$ is the eigenvalue matrix with decreasing magnitude, i.e., $\lambda_1^{(k)} \geq \cdots \geq \lambda_d^{(k)}$. Moreover, for $i \in \mathcal{S}$, vector $\mathbf{x}_i$ can be represented as a sum of two mutually orthogonal vectors $\mathbf{x}_i = \widehat{\mathbf{x}}_i + \tilde{\mathbf{x}}_i$, where $\widehat{\mathbf{x}}_i$ is the projection of vector $\mathbf{x}_i$ on the $r^{(k)}$-dimensional subspace spanned by the first $r^{(k)}$ eigenvalues, i.e., $\lambda_1^{(k)}, \dots, \lambda_{r^{(k)}}^{(k)}$, and $\tilde{\mathbf{x}}_i$ is the projection on the orthogonal subspace spanned by the remaining eigenvalues.

We take $\mathbf{U}^{(k)} = \{\mathbf{e}_1^{(k)}, \dots, \mathbf{e}_{r^{(k)}}^{(k)}\}$ as an estimate of the class-independent $r^{(k)}$-dimensional subspace with $r^{(k)} < d$ and

$$
r^{(k)} = \min\left\{ r^{(k)} : \sum_{i=1}^{r^{(k)}} \lambda_i^{(k)} \geq \sum_{i=1}^d \lambda_i^{(k)} \times \tau \right\} \quad (11)
$$

where $0 \leq \tau \leq 1$ is a threshold parameter controlling the loss of spectral information after projecting the data into the subspace.

2) Learning the MLR Regressors: In order to cope with difficulties in learning the regression vector $\boldsymbol{\omega}$ associated with bad or ill conditioning of the underlying inverse problem, we adopt a quadratic prior on $\boldsymbol{\omega}$ so that

$$
p(\boldsymbol{\omega}) \propto e^{-\beta/2\|\boldsymbol{\omega}\|^2} \quad (12)
$$

where $\beta \geq 0$ is a regularization parameter controlling weight of the prior.

In the present problem, learning the class densities amounts to estimating the logistic regressors $\boldsymbol{\omega}$. Inspired by previous works [7], [8], [18], [22], [30], we compute $\boldsymbol{\omega}$ by calculating the MAP estimate

$$
\widehat{\boldsymbol{\omega}} = \arg\max_{\boldsymbol{\omega}} \ell(\boldsymbol{\omega}) + \log p(\boldsymbol{\omega}) \quad (13)
$$

where $\ell(\boldsymbol{\omega})$ is the log-likelihood function given by

$$
\ell(\boldsymbol{\omega}) \equiv \log \prod_{i=1}^l p(y_i|\mathbf{x}_i, \boldsymbol{\omega}). \quad (14)
$$

The optimization problem in (13) is concave, although the term $\ell(\boldsymbol{\omega})$ is nonquadratic. This term can be approximated by a quadratic lower bound given by [7]; for any $k \in \mathcal{K}$ and the regressors $\boldsymbol{\omega}_t$ at iteration t, we have

$$
\begin{aligned}
\ell\left(\boldsymbol{\omega}^{(k)} \right) \geq \ell\left(\boldsymbol{\omega}_t^{(k)} \right) &+ \left(\boldsymbol{\omega}^{(k)} - \boldsymbol{\omega}_t^{(k)} \right)^{\mathrm{T}}\mathbf{g}\left(\boldsymbol{\omega}_t^{(k)} \right) \\
&+ \frac{1}{2}\left(\boldsymbol{\omega}^{(k)} - \boldsymbol{\omega}_t^{(k)} \right)^{\mathrm{T}}\mathbf{B}^{(k)}\left(\boldsymbol{\omega}^{(k)} - \boldsymbol{\omega}_t^{(k)} \right) \quad (15)
\end{aligned}
$$

with

$$
\mathbf{B}^{(k)} \equiv -(1/2)\left[\mathbf{I} - \mathbf{1}\mathbf{1}^{\mathrm{T}}/(K+1) \right] \otimes \sum_{i=1}^l \boldsymbol{\phi}^{(k)}(\mathbf{x}_i)\boldsymbol{\phi}^{(k)}(\mathbf{x}_i)^{\mathrm{T}} \quad (16)
$$

where $\mathbf{1}$ denotes a column vector of ones and $\mathbf{g}(\boldsymbol{\omega}_t^{(k)})$ is the gradient of $\ell(\cdot)$ at $\boldsymbol{\omega}_t^{(k)}$. Based on the lower bound (15), we implement an instance of the minorization–maximization (MM) algorithm [31], which consists in replacing, in each iteration, the objective function $\ell(\boldsymbol{\omega})$ with the lower bound (15) and then in maximizing it. It should be noted that the well-known expectation–maximization (EM) algorithm [32] is, in fact, an MM algorithm and that not all MM algorithms are EM instances. The aforementioned procedure then leads to

$$
\begin{aligned}
\widehat{\boldsymbol{\omega}}_{t+1}^{(k)} = \arg\max_{\boldsymbol{\omega}^{(k)}} \boldsymbol{\omega}^{(k)\mathrm{T}}&\left(\mathbf{g}\left(\widehat{\boldsymbol{\omega}}_t^{(k)} \right) - \mathbf{B}^{(k)}\widehat{\boldsymbol{\omega}}_t^{(k)} \right) \\
&+ \frac{1}{2}\boldsymbol{\omega}^{(k)\mathrm{T}}\left(\mathbf{B}^{(k)} - \beta\mathbf{I} \right)\boldsymbol{\omega}^{(k)}. \quad (17)
\end{aligned}
$$

Now, the optimization problem in (17) is quadratic and easy to solve, leading to the following update function:

$$\widehat{\boldsymbol{\omega}}_{t+1}^{(k)} = \frac{1}{2}\left(\mathbf{B}^{(k)} - \beta\mathbf{I}\right)^{-1}\left(\mathbf{B}^{(k)}\widehat{\boldsymbol{\omega}}_t^{(k)} - \mathbf{g}\left(\widehat{\boldsymbol{\omega}}_t^{(k)}\right)\right),$$

$$\text{for } k \in \mathcal{K}. \quad (18)$$

We note that matrix $(\mathbf{B}^{(k)} - \beta\mathbf{I})^{-1}$ is negative definite and, thus, nonsingular. Furthermore, it is fixed along the algorithm iterations; thus, it can be precomputed. With this in mind, it is now possible to perform an exact MAP-based MLR under a quadratic prior. The pseudocode for the subspace projection-based MLR algorithm, referred to hereinafter as MLR*sub*, is shown in Algorithm 1. In the algorithm description, *iters* denotes the maximum number of iterations. The overall complexity of Algorithm 1 is dominated by the computation of the correlation matrix, which has complexity $O(ld^2)$ (recall that l is the number of labeled samples and d is the dimensionality of the feature vectors).

Algorithm 1 MLR*sub*

Input: $\boldsymbol{\omega}_0, \mathcal{D}_l, \beta, \tau, iters$
Output: $\boldsymbol{\omega}, \mathbf{U} \equiv \{\mathbf{U}^{(1)}, \dots, \mathbf{U}^{(k)}\}$
for $k = 1$ to K **do**
    $\mathbf{U}^{(k)} \equiv \mathcal{U}(\mathcal{X}_{l(k)}^{(k)}, \tau)$ (* $\mathcal{U}$ computes the subspace according to (10)*)
    $\mathbf{B}^{(k)} \equiv \mathcal{B}(\mathbf{U}^{(k)}, \mathcal{D}_l)$ (* $\mathcal{B}$ computes the system matrix $\mathbf{B}$ according to (16)*)
end for
$t := 1$
while $t \le iters$ **or** stopping criterion is not satisfied **do**
    for $k := 1$ to K **do**
        $\mathbf{g}(\boldsymbol{\omega}_{t-1}^{(k)}) \equiv \nabla\ell(\boldsymbol{\omega}_{t-1}^{(k)})$
        $\boldsymbol{\omega}_t^{(k)} = \text{solution}\{\mathbf{B}^{(k)}, \mathbf{g}(\boldsymbol{\omega}_{t-1}^{(k)}), \mathbf{U}^{(k)}, \beta\}$
    end for
end while

B. MRF-Based MLL Spatial Prior

In order to improve the classification performance achieved by using the spectral information alone, in this paper, we integrate the contextual information with spectral information by using an isotropic MLL prior to modeling the image of class labels $\mathbf{y}$. This approach exploits the fact that, in segmenting real-world images, it is very likely that spatially neighboring pixels belong to the same class. This prior, which belongs to the MRF class, encourages piecewise smooth segmentations and promotes solutions in which adjacent pixels are likely to belong to the same class. The MLL prior constitutes a generalization of the Ising model [33] and has been widely used in image segmentation problems [34].

According to the Hammersly–Clifford theorem [35], the density associated with an MRF is a Gibbs' distribution [33].

Therefore, the prior model for segmentation has the following structure:

$$p(\mathbf{y}) = \frac{1}{Z}e^{\left(-\sum_{c \in \mathcal{C}} V_c(\mathbf{y})\right)} \quad (19)$$

where Z is a normalizing constant for the density, the sum in the exponent is over the so-called prior potentials $V_c(\mathbf{y})$ for the set of cliques[2] $\mathcal{C}$ over the image, and

$$-V_c(\mathbf{y}) = \begin{cases} \upsilon_{y_i}, & \text{if } |c| = 1 \text{ (single clique)} \\ \mu_c, & \text{if } |c| > 1 \text{ and } \forall_{i,j \in c} y_i = y_j \\ -\mu_c, & \text{if } |c| > 1 \text{ and } \exists_{i,j \in c} y_i \ne y_j \end{cases} \quad (20)$$

where μ_c is a nonnegative constant. The potential function in (20) encourages neighbors to have the same label. The introduced MLL prior offers a great deal of flexibility by allowing variations of the set of cliques and the parameters υ_{y_i} and μ_c. For example, the model generates texturelike regions if μ_c depends on c and bloblike regions otherwise [34]. In this paper, we take $\upsilon_{y_i} = c^{te}$ and $\mu_c = (1/2)\mu > 0$. Thus, (19) can be rewritten as follows:

$$p(\mathbf{y}) = \frac{1}{Z}e^{\mu \sum_{(i,j) \in \mathcal{C}} \delta(y_i - y_j)} \quad (21)$$

where $\delta(y)$ is the unit impulse function.[3] This choice gives no preference to any direction concerning. A straightforward computation of $p(y_i)$, i.e., the marginal of $p(\mathbf{y})$ with respect to y_i, leads to $p(y_i)$ being constant and, thus, equiprobable, which is therefore compatible with the assumption made in (2) and (8). Notice that the pairwise interaction terms $\delta(y_i - y_j)$ attach higher probability to equal neighboring labels than the other way around. In this way, the MLL prior promotes piecewise smooth segmentations, where parameter μ controls the level of smoothness.

C. MAP Estimate via Graph Cuts

Let us assume that the posterior class densities $p(y_i|\mathbf{x}_i)$ are estimated using (8). Let us also assume that the MLL prior $p(\mathbf{y})$ is estimated using (21). According to (3), the MAP segmentation is finally given by

$$\widehat{\mathbf{y}} = \arg\min_{\mathbf{y} \in \mathcal{K}^n}\left\{\sum_{i \in \mathcal{S}} -\log p(y_i|\mathbf{x}_i, \widehat{\boldsymbol{\omega}}) - \mu\sum_{i \sim j}\delta(y_i - y_j)\right\}. \quad (22)$$

This is a combinatorial optimization problem involving unary and pairwise interaction terms, which is very difficult to compute. Several new algorithms such as graph cuts [25], [36], [37], loopy belief propagation [38], [39], and tree-reweighted message passing [40] have been proposed in the literature in order to tackle this optimization problem. In this paper, we resort to the α-Expansion graph-cut-based algorithm [25], [41].

---

[2]A clique is a single term or either a set of pixels that are neighbors of one another.
[3]That is, $\delta(0) = 1$ and $\delta(y) = 0$ for $y \ne 0$.
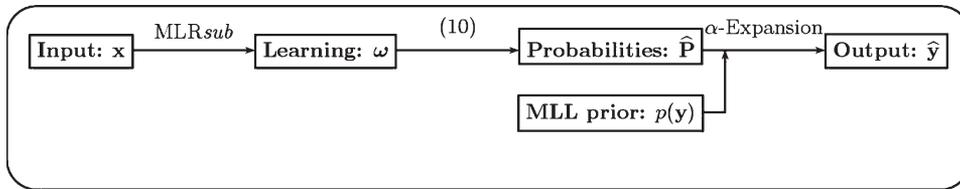
Fig. 2.   Block diagram summarizing the most relevant steps of the proposed MLR*sub*MLL algorithm.

This method yields good approximations to the MAP segmentation and is quite efficient from a computational viewpoint, with practical computational complexity $O(n)$ [25].

D. Supervised Segmentation Algorithm: MLRsubMLL

To conclude the description of our proposed method, Algorithm 2 provides a pseudocode for our newly developed supervised segmentation algorithm based on a subspace MLR classifier with MRF-based MLL prior. This algorithm, called MLR*sub*MLL hereinafter, integrates all the different modules described in this section. Specifically, line 3 in Algorithm 2 learns the logistic regressors using MLR*sub*, which is applied to the full hyperspectral image. Here, the quadratic regularization parameter $\beta \geq 0$ is used to tackle ill-conditioned problems. Line 4 in Algorithm 2 computes the probabilities based on the outcome of MLR*sub*. Line 5 in Algorithm 2 efficiently computes the MAP segmentation by applying the α-Expansion graph-cut-based algorithm, where the neighborhood parameter μ determines the strength of the spatial prior. For illustrative purposes, Fig. 2 summarizes the most relevant steps of the newly proposed segmentation algorithm using a block diagram.

Algorithm 2 MLR*sub*MLL

1: **Input**: $\mathbf{x}, \mathcal{D}_l, \beta, \tau, \mu$
2: **Output**: $\widehat{\mathbf{y}}$
3: $\{\widehat{\boldsymbol{\omega}}, \mathbf{U}\} = \text{MLR}sub\{\mathcal{D}_l, \beta, \tau\}$
4: $\widehat{\mathbf{P}} := \widehat{\mathbf{p}}(\mathbf{x}, \widehat{\boldsymbol{\omega}}, \mathbf{U})$ ($* \widehat{\mathbf{P}}$ collects the probabilities in (9)$*$)
5: $\widehat{\mathbf{y}} := \alpha - \text{Expansion}(\widehat{\mathbf{P}}, \mu, \text{neighborhood})$

The overall complexity of the proposed MLR*sub*MLL algorithm is dominated by the MLR*sub* algorithm inferring the regressors, which has computational complexity $O(ld^2)$, and also by the α-Expansion algorithm used to determine the MAP segmentation, which has practical complexity $O(n)$. In conclusion, if $ld^2 > n$ (e.g., the problem is high dimensional, with a large number of training samples), then the overall complexity is dominated by the subspace-based learning step. Otherwise, if $ld^2 < n$ (e.g., the problem is given by a large number of pixels), then the overall complexity is dominated by the α-Expansion algorithm.

IV. Experimental Results

This section uses both simulated and real hyperspectral data sets to illustrate the effectiveness of the proposed MLR*sub*MLL

segmentation algorithm in different analysis scenarios. The main goal of using simulated data sets is to assess the performance of the algorithm in a fully controlled environment, whereas the main goal of using real data sets is to compare the algorithm with other state-of-the-art analysis techniques using widely used hyperspectral scenes. The remainder of this section is organized as follows. Section IV-A first explains the parameter settings adopted in our experimental evaluation. Section IV-B then evaluates the proposed MLR*sub*MLL algorithm by using simulated data sets, whereas Section IV-C evaluates the proposed segmentation algorithm using real hyperspectral images.

A. Parameter Settings

Before describing our results with simulated and real hyperspectral data sets, it is first important to discuss the parameter settings adopted in our experiments. In our tests, we assume $l^{(k)} \simeq l/K$ for $k \in \mathcal{K}$. For small classes, if the total number of labeled samples per class k in the ground-truth image, for example, $L^{(k)}$, is smaller than l/K, we take $l^{(k)} = L^{(k)}/2$. In this case, we use more labeled samples to represent large classes. It should be noted that, in all experiments, the labeled sets $\mathcal{D}_l$ are randomly selected from the available labeled samples and that the remaining samples are used for validation. Each value of overall accuracy [OA (in percent)] is obtained after conducting ten Monte Carlo runs with respect to the labeled samples $\mathcal{D}_l$. The labeled samples for each Monte Carlo simulation are obtained by resampling the available labeled samples. Prior to the experiments, we infer the setting of the quadratic parameter β. In practice, β is relevant to the condition number of $\mathbf{B}^{(k)}$ for $k \in \mathcal{K}$. In this paper, we set $\beta = e^{-10}$ for all experiments.

B. Experiments With Simulated Hyperspectral Data

In our experiments, we have generated a simulated hyperspectral scene as follows. First, we generate an image of features using a linear mixture model

$$\mathbf{x}_i = \sum_{k=1}^{K} \mathbf{m}^{(k)} \gamma_i^{(k)} + \mathbf{n}_i \tag{23}$$

with $K = 10$. Here, $\mathbf{m}^{(k)}$ for $k \in \mathcal{K}$ are spectral signatures obtained from the U.S. Geological Survey (USGS) digital spectral library,[4] and $\mathbf{x}_i$ is a simulated mixed pixel. An MLL distribution with smoothness parameter $\mu = 2$ is used to generate the spatial information, and the total size of the simulated image is of

---

[4]The USGS library of spectral signatures is available online: http://speclab. cr.usgs.gov.

(a) Ground truth   (b) Classification map   (c) Segmentation map
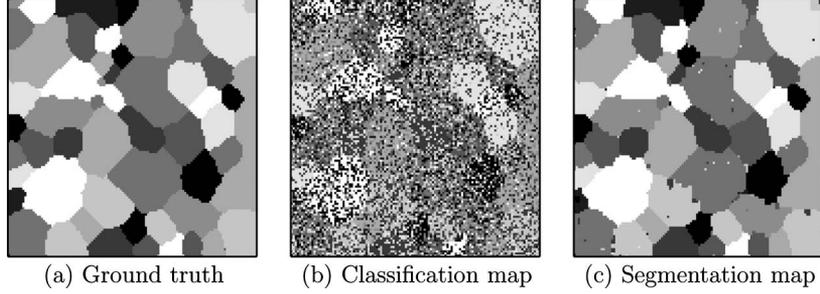
Fig. 3. Classification and segmentation maps obtained after applying the proposed method to a simulated hyperspectral scene with $\sigma = 0.8$ and $\gamma = 0.7$ by using $\tau = 0.9$, $l = 288$, and $\mu = 2$. (a) Ground-truth class labels. (b) Classification result (OA $= 49.09\%$). (c) Segmentation result (OA $= 94.34\%$).

120×120 pixels. Zero-mean Gaussian noise with covariance $\sigma^2\mathbf{I}$, i.e., $\mathbf{n}_i \sim \mathcal{N}(0, \sigma^2\mathbf{I})$ is finally added to our simple simulated hyperspectral scene. For illustrative purposes, the image of class labels $\mathbf{y}$ is shown in Fig. 3(a). Assume that $\mathbf{x}_i$ has class label $y_i = k_k$; then, we define $\gamma_i^{(k_k)}$ as the abundance of the objective class and $\gamma_i^{(k)}$ (for $k \in \mathcal{K}$ and $k \neq k_k$) as the abundance of the remaining signatures which contribute to the mixed pixel, where $\gamma_i^{(k)}$ values are generated according to a simple uniform distribution in the proposed problem. In order to simplify notations, we take $\gamma_i^{(k_k)} = \gamma$, $\sum_{k \in \mathcal{K}, k \neq k_i} \gamma_i^{(k)} = 1 - \gamma$, and we use the same γ for all pixels.

We have conducted five different experiments with the simulated hyperspectral image described earlier. These experiments have been carefully designed in order to analyze several relevant aspects of our proposed MLR*sub*MLL segmentation algorithm in a fully controlled environment.

1) In our first experiment, we evaluate the impact of the presence of mixed pixels on the segmentation output.
2) In our second experiment, we analyze the impact of the parameter τ (controlling the amount of spectral information retained after subspace projection) on the segmentation output.
3) In our third experiment, we evaluate the impact of the training set size on the segmentation output.
4) In our fourth experiment, we analyze the impact of the smoothness parameter μ on the segmentation output.
5) In our fifth experiment, we evaluate the impact of noise on the segmentation output.

In all these experiments, we will use the optimal value of classification accuracy (OA$_{\text{opt}}$) as a reference to evaluate the goodness of our reported OA scores. Here, OA$_{\text{opt}} \equiv 100(1 - P_e)\%$, where P_e is defined as follows [42]:

$$P_e = p(y_i \neq \widehat{y}_i) \qquad (24)$$

where y_i and $\widehat{y}_i$ are the true label and the MAP estimate, respectively, i.e.,

$$\widehat{y}_i = \arg\max_{y_i} p(y_i|\mathbf{x}_i).$$

For a multiclass problem, we use the following error bound as an alternative since (24) is difficult to compute:

$$\text{erfc}\left(\frac{dist_{\min}}{2\sigma}\right) \leq P_e \leq \frac{K-1}{2}\text{erfc}\left(\frac{dist_{\min}}{2\sigma}\right) \qquad (25)$$
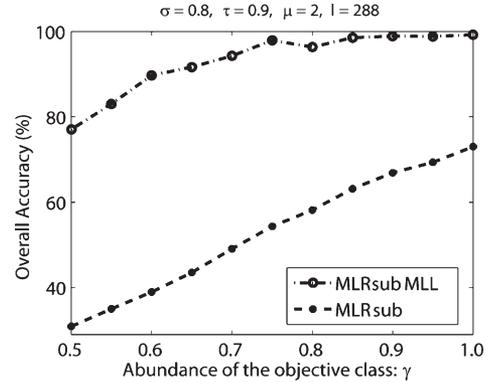
Fig. 4. OA results as a function of the abundance of the objective class: γ, with $\tau = 0.9$, $\mu = 2$, and $l = 288$ for a problem with mixed pixels and $\sigma = 0.8$. Dash–dot lines with circles denote the segmentation results obtained by the MLR*sub*MLL algorithm; dashed lines with asterisks denote the classification results obtained by the MLR*sub* algorithm.

where $\text{erfc}(\cdot)$ denotes the complementary error function and $dist_{\min}$ denotes the minimum distance between any point of mean vectors, i.e., $dist_{\min} = \min_{i \neq j} \|\mathbf{m}_i - \mathbf{m}_j\|$ for any $i, j \in \mathcal{K}$. This is the so-called union bound [12], which is widely used in multiclass problems. However, union bound is not a good measurement to present the difficulty because of the mixtures. Nevertheless, it is worth noting that as γ decreases, the difficulty increases, i.e., OA$_{\text{opt}}$ decreases. Thus, in this paper, we use the union bound, while $\gamma = 1$, to define the difficulty of our problem.

1) Experiment 1—Impact of the Presence of Mixed Pixels: In this experiment, we first consider a problem with $\sigma = 0.8$ by using $\tau = 0.9$, $\mu = 2$, and $\gamma \in [0.5 \ 1.0]$. In this context, the optimal value of classification accuracy is given by OA$_{\text{opt}} \leq 71.04\%$ with $\gamma = 1$. It should be noted that the values of parameters τ and μ in our simulation are probably suboptimal. However, we have decided to fix them to the specified values because we have experimentally observed that these settings lead to good performance in the considered analysis scenario. Fig. 4 shows the obtained OA results as a function of γ (which determines the degree of spectral purity in the simulated pixels). In order to show the good capability of the proposed MLR*sub*MLL in the task of dealing with limited training sets, only 288 labeled samples (2% of the available samples, evenly distributed among classes) are used as the training set. Notice the good performance achieved by the proposed MLR*sub*MLL algorithm with the classes dominated by mixed pixels. In those classes, the segmentation results provided by MLR*sub*MLL

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

8

IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING
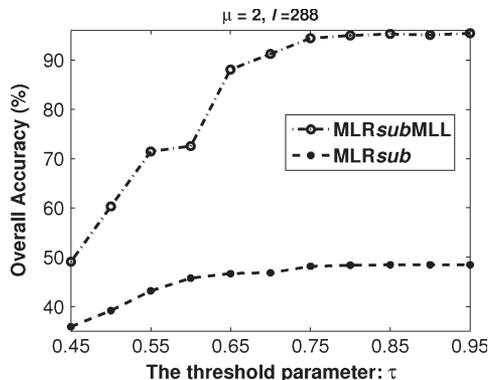
Fig. 5. OA results (as a function of τ) with $\mu = 2$ for a problem with $\sigma = 0.8$ and $\gamma = 0.7$.

significantly outperform the classification results obtained by the MLR*sub* using only the spectral information. For illustrative purposes, Fig. 3(b) and (c) shows the respective classification and segmentation maps obtained for the problem with $\sigma = 0.8$ and $\gamma = 0.7$, using $\tau = 0.9$, $\mu = 2$, and $l = 288$. Moreover, Fig. 4 shows that the performances of both MLR*sub*MLL and MLR*sub* increase as the abundance of the objective classes increase. This is expected since the problem is easier to solve as the presence of mixed pixels is decreased. In the following experiments, we will consider $\gamma = 0.7$ which leads to a difficult segmentation problem, as shown in Fig. 4.

2) Experiment 2—Impact of Parameter τ: In our second experiment, we analyze the impact of the threshold parameter τ intended to control the loss of spectral information after projecting the original hyperspectral data into a subspace. This parameter is directly related with the number of components retained after the projection and with the amount of information comprised by the retained components. To address this issue, we analyze the performance of the proposed methods for different values of τ in a problem with $\sigma = 0.8$ (OA$_{\text{opt}} \leq 71.04\%$ with $\gamma = 1$) and $\gamma = 0.7$ by using $\mu = 2$. Fig. 5 shows the OAs obtained by the proposed MLR*sub* and MLR*sub*MLL algorithms as a function of τ, where 288 labeled samples are again used as the (limited) training set. Notice the good performance achieved by the proposed MLR*sub*MLL segmentation algorithm, which yielded much better OA results than OA$_{\text{opt}}$ ($\leq 71.04\%$) in cases with $\tau \geq 0.65$. Furthermore, both classification and segmentation results increase as τ increases. This is reasonable since the amount of spectral information that is retained after the projection of the original data into the subspace is increased as τ increases. This also indicates that the proposed methods can perform well in the presence of limited training sets, even after the dimensionality of the subspace is increased. The robustness of the proposed methods in the presence of very limited training sets is analyzed in more detail in the following experiment.

3) Experiment 3—Impact of the Training Set Size: In our third simulated image experiment, we analyze the impact of the training set size on the segmentation performance. Fig. 6(a) and (b) shows the OA and standard deviation (std) results, respectively, obtained by our proposed methods as a function of the number of labeled samples (l) used in the training process with $\tau = 0.9$ and $\mu = 2$. Again, these parameter settings may

be suboptimal but lead to very good results in our experiments. Notice the quality of the segmentation results obtained by our proposed MLR*sub*MLL algorithm, which shows high robustness even with very limited training set sizes. As the number of labeled samples increases, the OA increases and the standard deviation decreases. This is expected since an increase of the number of labeled samples should decrease in the uncertainty when estimating the subspace for each class.

On the other hand, we have experimentally observed that the OA and the standard deviation results converge to very high and very low values, respectively, for a certain number of labeled samples. In our particular case, the use of 350 labeled samples resulted in an OA of 97.76% with std = 0.37. This indicates that robust generalization can be achieved by the combination of subspace-based MLR regressors and spatial–contextual information. From this experiment, we can conclude that our proposed algorithm converges to almost identical results once the class-independent subspaces are well estimated using a *sufficient* number of labeled training samples, where the term *sufficient* in our experiments means a low percentage of labeled samples. This is because, as already mentioned, the classes normally live in a very low dimensional subspace. Despite the encouraging results obtained thus far with the conducted simulations, a more detailed investigation of two additional aspects, i.e., the relevance of the smoothness parameter μ on spatial characterization and the overall performance of our proposed approaches in the presence of different noise levels, should be conducted. This will be done in the next two experiments performed with our simulated hyperspectral scene.

4) Experiment 4—Impact of Parameter μ: In this experiment, we conduct an evaluation of the impact of the smoothness parameter μ on the obtained segmentation results. In practice, we use the cross-validation sampling method [43] to estimate μ by using available training samples. Fig. 7 plots the obtained OA results as a function of μ, with $\tau = 0.9$ and $l = 288$ (2% of the available samples per class evenly distributed among classes). From Fig. 7, we can conclude that the segmentation performance indeed depends on the setting of μ. However, even with a suboptimal parameter setting $1.5 \leq \mu \leq 4$, the proposed MLR*sub*MLL algorithm leads to good segmentation results for the considered problem. This indicates that the algorithm is not very sensitive to the setting of parameter μ since all values of this parameter in a certain range of interest ultimately lead to high values of the OA for the considered problem.

It should be noted that, in all experiments conducted thus far, the noise standard deviation considered in the simulations was $\sigma = 0.8$ (a reasonable parameter setting according to our tests). However, a remaining aspect to be analyzed is the sensitivity of the proposed method to different noise levels.

5) Experiment 5—Impact of Noise: In our last experiment with simulated data, we evaluate the impact of noise on the proposed segmentation algorithm by using only $l = 288$ labeled samples (2% of the available samples per class evenly distributed among classes) as in previous experiments. Fig. 8 shows the OA results as a function of the noise standard deviation σ for two different problems: (a) $\gamma = 1$ and (b) $\gamma = 0.7$. As shown in Fig. 8, the performance of the proposed MLR*sub*MLL algorithm decreases as σ increases, but the increase in the OAs

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

LI *et al.*: SPECTRAL–SPATIAL HYPERSPECTRAL IMAGE SEGMENTATION USING SUBSPACE MLR AND MRFS   9
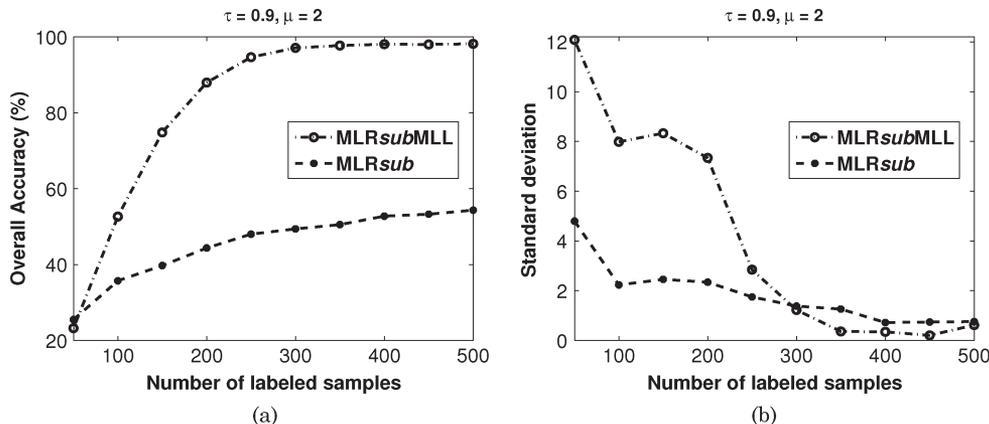
Fig. 6. Classification and segmentation results obtained for a problem with $\sigma = 0.8$ and $\gamma = 0.7$ by using $\tau = 0.9$ and $\mu = 2$. (a) OA results as a function of the number of labeled samples. (b) Standard deviation results as a function of the number of labeled samples.
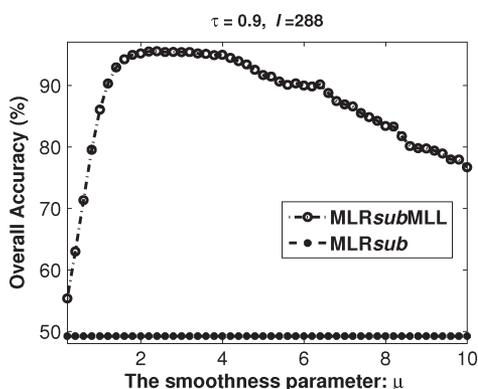
Fig. 7. OA results as a function of the smoothness parameter μ for a problem with $\sigma = 0.8$ and $\gamma = 0.7$ with $\tau = 0.9$ and $l = 288$.

obtained with regard to the MLR*sub* classification are always remarkable. From Fig. 8, we can also conclude that the results achieved by our proposed segmentation algorithm are superior to the OA$_{\text{opt}}$ result. Specifically, for the problem with $\sigma = 1.5$, the MLR*sub*MLL obtained a segmentation OA of 58.12% with $\gamma = 0.7$ [see Fig. 8(b)], which is 15.34% higher than the optimal value (OA$_{\text{opt}} \leq 42.78\%$ with $\gamma = 1$) in Fig. 8(a).

Summarizing, the experimental results conducted with simulated data sets indicate that the proposed MLR*sub*MLL algorithm achieves adequate performance in highly mixed and noisy environments and with limited training sets, exhibiting robustness for a wide range of parameter settings that simplify the choice of such parameters by the end user. In other words, although the performance of the algorithm has been shown to be dependent on the setting of parameters τ and μ, suboptimal settings of these parameters are easy to obtain and lead to good characterization results in different simulation environments. Although the experimental evaluation conducted with simulated data sets provided very encouraging results, further analyses with real hyperspectral scenes and comparisons with other state-of-the-art methods are highly desirable in order to fully substantiate the proposed method.

C. Experiments With Real Hyperspectral Data

In order to evaluate the proposed MLR*sub*MLL algorithm in real analysis scenarios, we use two widely used hyperspec-

tral data sets collected by AVIRIS and the Reflective Optics System Spectrographic Imaging System (ROSIS), respectively, operated by the German Aerospace Agency (DLR). For the purpose of comparison, we use other state-of-the-art supervised classifiers such as linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), logistic discriminant analysis (LogDA), and SVMs [12], [44], [45], which are well-established techniques in the machine learning community [6], [46]–[48]. For these methods, we project the original hyperspectral data sets into a subspace by using the hyperspectral signal identification by minimum error (HySime) method [49] which was observed to perform better than standard eigenvector calculation considered for the other tested methods; hence, we decided to report the best obtained results for each considered method. In all cases, the loss of spectral information after projecting the data into the subspace was controlled by parameter τ.

Furthermore, in order to have a fair comparison with our segmentation method (which includes spatial–contextual information), in this paper, we have also expanded the considered discriminant analysis approaches (LDA, QDA, and LogDA) with the MLL spatial prior to obtain segmentation methods (referred to hereinafter as LDAMLL, QDAMLL, and Log-DAMLL) that can be compared with our proposed algorithm. In all experiments, we empirically set $\tau = 0.999$ and $\mu = 2$. Although suboptimal, we have experimentally tested that these settings lead to good characterization results with all the considered data sets, a fact that reveals that the proposed approach can perform accurately using a variety of hyperspectral scenes collected by different instruments. Regarding parameter τ, we have also experimentally analyzed its influence in the final segmentation and observed that high-quality segmentations can be already obtained for values of $\tau > 0.75$ as it was already the case in our simulated image experiments. Hence, we conclude that the impact of this parameter on the final segmentation is not very significant.

1) Experiment 1—AVIRIS Indian Pines Data Set: In our first experiment, we use the well-known AVIRIS Indian Pines data set to analyze the performance of the proposed algorithm in comparison with other methods. The scene contains 145×145 pixels and 202 spectral bands. The ground-truth data contains 16 mutually exclusive classes and a total of 10 366

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

10   IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING
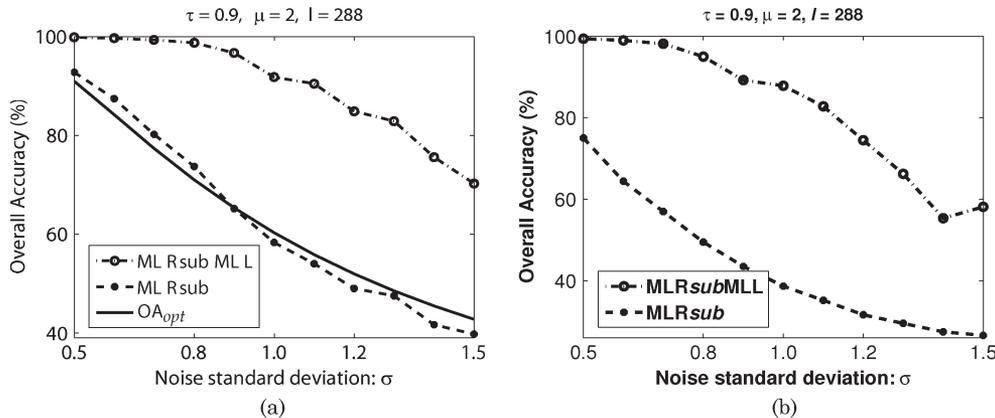
Fig. 8. OA results achieved (for two different values of γ) as a function of the noise standard deviation σ with $\tau = 0.9$, $\mu = 2$, and $l = 288$. (a) $\gamma = 1$. (b) $\gamma = 0.7$.

TABLE I
OVERALL, AVERAGE, AND INDIVIDUAL CLASS ACCURACIES (IN PERCENT) AND κ STATISTIC OBTAINED FOR THE AVIRIS INDIAN
PINES DATA SET. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD TYPEFACE

Class	# samples	Classification algorithms					Segmentation algorithms			
		MLR*sub*	LDA	QDA	LogDA	SVM	MLR*sub*MLL	LDAMLL	QDAMLL	LogDAMLL
Alfalfa	54	66.94	82.34	55.91	83.50	94.48	94.62	95.38	67.94	**97.69**
Corn-no till	1434	59.05	64.97	68.88	64.42	71.34	86.51	76.37	81.95	**96.12**
Corn-min till	834	54.48	41.43	71.05	60.72	68.00	87.07	51.29	**87.27**	81.14
Corn	234	73.68	68.76	88.65	77.04	85.23	91.72	78.16	92.42	**93.16**
Grass/pasture	497	57.43	68.89	73.37	67.68	73.19	**86.27**	82.44	82.38	79.11
Grass/tree	747	98.90	97.31	96.82	94.77	96.94	**99.52**	99.76	98.43	99.12
Grass/pasture-mowed	26	91.01	53.88	74.65	68.97	77.77	85.54	71.99	**87.40**	85.97
Hay-windrowed	489	71.89	87.00	19.56	76.67	85.89	90.00	**92.00**	23.67	90.00
Oats	20	71.41	48.60	70.42	67.90	73.74	**97.88**	77.28	89.69	90.23
Soybeans-no till	968	75.13	67.25	86.82	75.66	86.31	**93.16**	73.79	95.16	83.01
Soybeans-min till	2468	77.25	66.76	86.58	74.45	87.03	98.35	94.44	97.30	**99.12**
Soybeans-clean till	614	90.47	70.35	89.86	89.22	92.71	96.42	69.44	86.49	**94.13**
Wheat	212	93.85	95.12	93.23	90.33	94.27	**99.18**	98.75	96.44	98.36
Woods	1294	91.64	84.99	91.64	95.37	97.42	**98.86**	85.58	98.57	**98.86**
Bldg-Grass-Tree-Drives	380	99.33	99.73	97.71	97.92	99.13	99.67	**99.87**	98.79	99.73
Stone-steel towers	95	94.40	84.32	94.68	92.13	90.29	**98.43**	87.30	95.78	93.68
OA		74.15	65.22	79.84	75.42	80.56	**93.66**	79.41	90.02	89.23
AA		77.30	73.67	78.73	79.82	77.81	**93.95**	83.36	86.23	90.46
κ		70.30	60.61	77.02	72.00	85.86	**92.69**	76.41	88.56	87.66

labeled pixels. This image is a classical benchmark to validate the accuracy of hyperspectral image analysis algorithms and constitutes a challenging problem due to the significant presence of mixed pixels in all available classes and also because of the unbalanced number of available labeled pixels per class.

In order to test the performance of the proposed algorithms with limited training sets, a total size of $l = 1036$ (which represents 10% of the available labeled samples evenly distributed among classes) is used for training purposes, where the remaining 90% of the samples were used for validation. Table I illustrates the OA, average accuracy (AA), kappa statistic coefficient (κ), and individual class accuracy (in percent) results achieved by the proposed algorithms after ten Monte Carlo runs. By adopting an MLL spatial prior, the segmentation algorithms significantly improved the classification results obtained by the considered classification algorithms. For instance, the MLR*sub*MLL obtained an OA of 93.66%, 19.51% larger than

that obtained by the MLL*sub* algorithm, whereas the QDAMLL obtained an OA of 90.02%, which is 10.18% higher than the result obtained by the QDA algorithm. It is remarkable that the MLR*sub* algorithm did not provide the best classification results in our experiments (it only outperformed the classification results provided by LDA). However, the inclusion of the MLL prior improved more significantly the results obtained by MLR*sub* than those obtained by the other discriminant analysis methods. This is because the proposed MLR*sub* method (enhanced with the inclusion of a subspace projection method) leads to very reliable class posterior probabilities for each class after reducing the negative effects caused by noise and mixed pixels. As a result, the obtained class posterior probabilities play essential roles in the complete posterior probabilities such that the final MAP segmentation can greatly benefit from the inclusion of both the spectral and the spatial information available in the data.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

LI *et al.*: SPECTRAL–SPATIAL HYPERSPECTRAL IMAGE SEGMENTATION USING SUBSPACE MLR AND MRFS

11
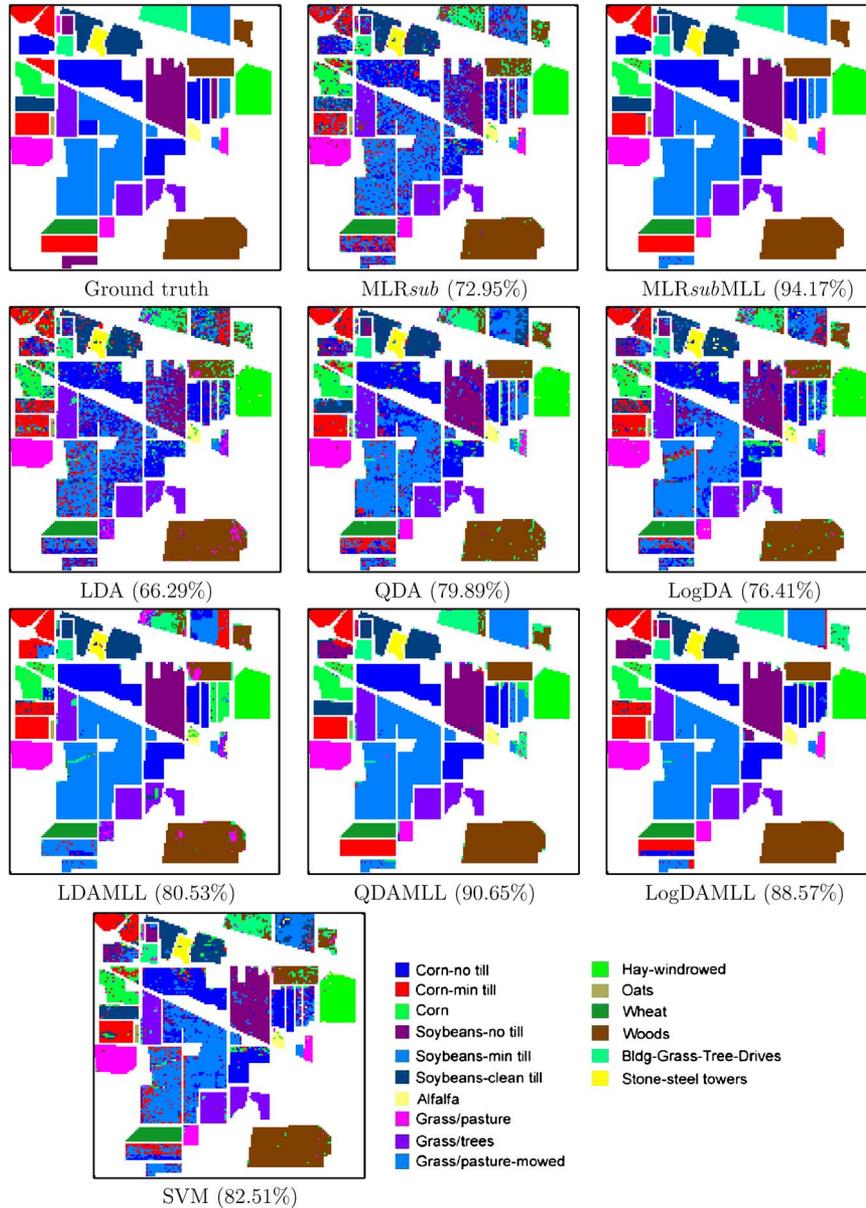
Fig. 9. Classification/segmentation maps obtained by the different tested methods for the AVIRIS Indian Pines scene (overall accuracies are reported in parentheses).

For illustrative purposes, Fig. 9 shows the ground truth and some of the classification/segmentation results obtained by the different tested methods for the AVIRIS Indian Pines scene. For each method, we randomly selected one of the maps obtained after conducting ten Monte Carlo runs. As shown by Fig. 9, the SVM produced the best classification map while the MLR*sub*MLL produced the best segmentation map. An immediate issue resulting from experiments in Fig. 9 is whether the use of spatial–contextual information could result in an increase in the SVM classification results. In order to analyze this issue in more detail, in the following experiment, we will consider a recently developed SVM-based classifier which combines spatial and spectral information [19]. Furthermore, we will also consider a segmentation method based on the watershed transform [21]. The results for these methods were only available to us in the framework of experiments previously conducted with

the ROSIS University of Pavia data set [9], [21] and hence could not be included in the AVIRIS Indian Pines image experiments.

2) Experiment 2— ROSIS University of Pavia Data Set: The second real hyperspectral data set that we have considered in experiments was acquired in 2001 by the ROSIS instrument, flown over the city of Pavia, Italy. The image scene, with size of 610×340 pixels, is centered at the University of Pavia. After removing 12 bands due to noise and water absorption, it comprises 103 spectral channels. Nine ground-truth classes, with a total of 3921 training samples and 42 776 test samples, were considered in experiments. Fig. 10(a) shows a false-color composite of the image, while Fig. 10(b) shows nine ground-truth classes of interest, which comprise urban features, as well as soil and vegetation features.

In our first test, we used the entire training set available for this scene in order to train different classifiers. Table II reports
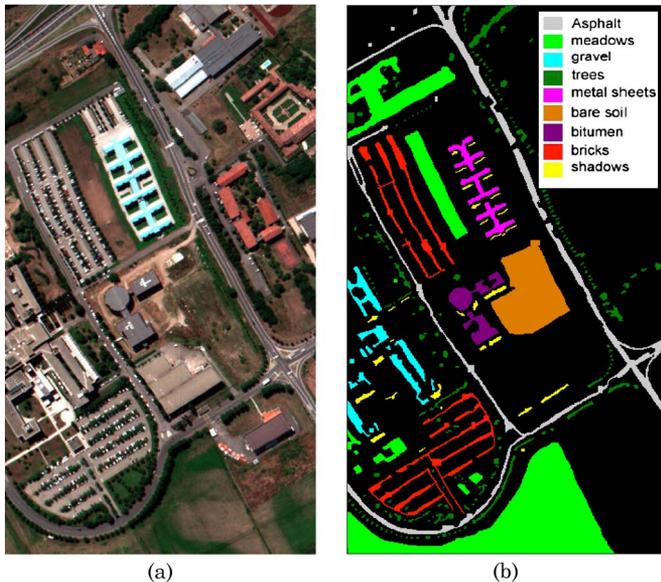
Fig. 10. (a) False-color composition of the ROSIS Pavia scene. (b) Ground-truth map containing nine mutually exclusive land-cover classes.

the obtained values of OA, AA, κ, and individual accuracies. In this comparison, we included the same set of classifiers used in the experiments with the AVIRIS Indian Pines image, along with two additional spectral–spatial classifiers: an SVM-based classifier trained with extended morphological profiles (designated in the table by EMP/SVM) [19] and a segmentation method based on the watershed transform [21]. The results reported in the table are respectively taken from [9] and [21], where exactly the same training and test sets aforementioned were used to produce the reported results, thus allowing a fair intercomparison of methods. By using the entire training set, the proposed MLRsubMLL algorithm obtained an OA of 94.10% in the considered analysis scenario. For illustrative purposes, Fig. 11 shows the classification and segmentation maps achieved by some of the considered methods.

Additionally, Table III provides a comparison of the proposed MLRsubMLL algorithm with other spatial–spectral methods such as the well-known extraction and classification of homogeneous objects algorithm in [1], other SVM- and MRF-based spectral-spatial segmentation algorithms presented in [50], and our previous work based on MLR spectral model and MLL prior in [22] for the ROSIS University of Pavia data set. The results presented in Table III are obtained using exactly the same training and test sets. From Table III, we can conclude that MLRsubMLL obtained the best results in terms of OA and κ and comparable results with those of the other tested algorithms in terms of AA. It should be noted that these results could be improved even more by merging the results obtained by different methods (i.e., by applying pixelwise majority voting [51]). Although not explicitly explored in this paper, this strategy represents a potential improvement that will be the subject of our future work.

In our second test, we analyze the sensitivity of the considered methods to different training sets made up of a limited number of samples. For this purpose, we constructed small training sets by randomly selecting 20, 30, 40, 60, 80, and

100 labeled samples per class from the original training set. Fig. 12 shows the obtained OA results by the different methods as a function of the number of labeled samples per class. By using only 60 labeled samples per class ($l = 540$ samples, which represents around 14% of the entire training set), the proposed MLRsubMLL obtained an OA of 88.85%. This result is quite remarkable since, for instance, the OA obtained by the EMP/SVM algorithm by using the entire training set was slightly lower (85.22%). When a spatial prior was adopted, the segmentation algorithms in Fig. 12(b) always achieved significantly better results than their classification counterparts in Fig. 12(a), thus indicating the importance of including spatial–contextual information.

A critical issue here is that, in Fig. 12(a), SVM provided very good results in comparison with the competitors. However, in Fig. 12(b), no segmentation results, such as SVM+MLL, are reported. In this paper, we did not implement SVM+MLL. This is because, as the SVM is a hard classifier, the class probabilities estimated by it are not reliable for estimating the complete MAP segmentation. Nevertheless, it is noticeable that, in Tables II and III, the proposed MLRsubMLL obtained very good results while compared with SVM- and EMP or MRF-based spectral-spatial-based algorithms.

V. CONCLUSION

In this paper, we have developed a new spectral-spatial segmentation approach which combines MLR with a subspace projection method to better characterize noise and mixed pixels. It includes contextual information using an MLL Markov–Gibbs prior. By computing the MAP segmentation with an optimized α-Expansion graph-cut-based algorithm, the proposed segmentation method provides good accuracies when compared with other methods. It also exhibits robustness to different criteria, such as noise, presence of mixed pixels, and limited availability of training samples without the need for fine-tuning of input parameters. Although our experimental results are competitive with those reported for other state-of-the-art spectral and spectral-spatial classification/segmentation methods, further work should be focused on conducting additional experiments with real hyperspectral scenes collected by other instruments, such as the new generation of spaceborne instruments that are currently under development. Given the similar spectral and spatial resolutions of these instruments with regard to the airborne systems adopted in our real experiments, we anticipate that the proposed robust segmentation techniques can also perform accurately with the new generation of satellite instruments. Another important research line deserving future experimentation focuses on the fusion/aggregation of the results obtained by different classifiers, i.e., by merging the results obtained by different methods using pixelwise majority voting.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

LI *et al.*: SPECTRAL–SPATIAL HYPERSPECTRAL IMAGE SEGMENTATION USING SUBSPACE MLR AND MRFS   13

TABLE II
OVERALL, AVERAGE, AND INDIVIDUAL CLASS ACCURACIES (IN PERCENT) AND κ STATISTIC OBTAINED FOR THE ROSIS UNIVERSITY OF PAVIA DATA SET. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD TYPEFACE

Class	# samples		Classification algorithms					Segmentation algorithms					
	Train	Test	MLR*sub*	LDA	QDA	LogDA	SVM†	MLR*sub*MLL	LDAMLL	QDAMLL	LogDAMLL	EMP/SVM†	[21]‡
Asphalt	548	6631	65.63	69.45	67.65	70.89	83.71	93.83	89.56	90.68	86.91	**95.36**	93.64
Bare soil	540	18649	64.80	46.59	73.49	75.06	92.25	98.43	45.93	98.27	**98.57**	63.72	97.35
Bitumen	392	2099	81.28	63.31	93.53	83.98	81.58	**99.32**	62.11	93.38	90.38	98.87	96.23
Bricks	524	3064	59.78	88.29	89.52	87.91	92.59	95.19	**99.08**	98.45	97.39	95.41	97.92
Gravel	265	1345	66.51	39.11	59.79	55.31	70.32	71.13	26.82	62.17	38.21	**87.61**	66.12
Meadows	532	5029	64.19	81.92	75.73	76.72	70.25	**94.80**	85.15	84.73	84.59	80.33	75.09
Metal sheets	375	1330	99.78	99.41	99.93	**100**	99.41	**100**	99.70	99.93	**100**	99.48	99.91
Shadows	514	3682	92.82	99.79	99.26	99.79	96.62	96.20	**100**	99.79	94.09	97.68	96.98
Trees	231	947	72.19	95.07	96.64	96.38	97.81	92.17	94.09	**99.80**	94.94	98.37	98.56
OA			67.08	75.59	77.95	78.41	80.99	**94.10**	80.27	89.48	87.04	85.22	85.42
AA			74.11	75.88	83.95	82.90	88.28	**93.45**	78.05	91.91	83.32	90.76	91.31
κ			58.53	68.16	71.93	72.47	76.16	**92.24**	73.90	86.46	87.23	80.86	81.30

Notes:
† Results are directly taken from [9], which used EMPs for spectral-spatial characterization prior to SVM-based classification.
‡ Results are directly taken from [21], which used a spectral-spatial classifier based on a pixel-wise SVM classifier with majority voting within the watershed regions to produce the final segmentation.
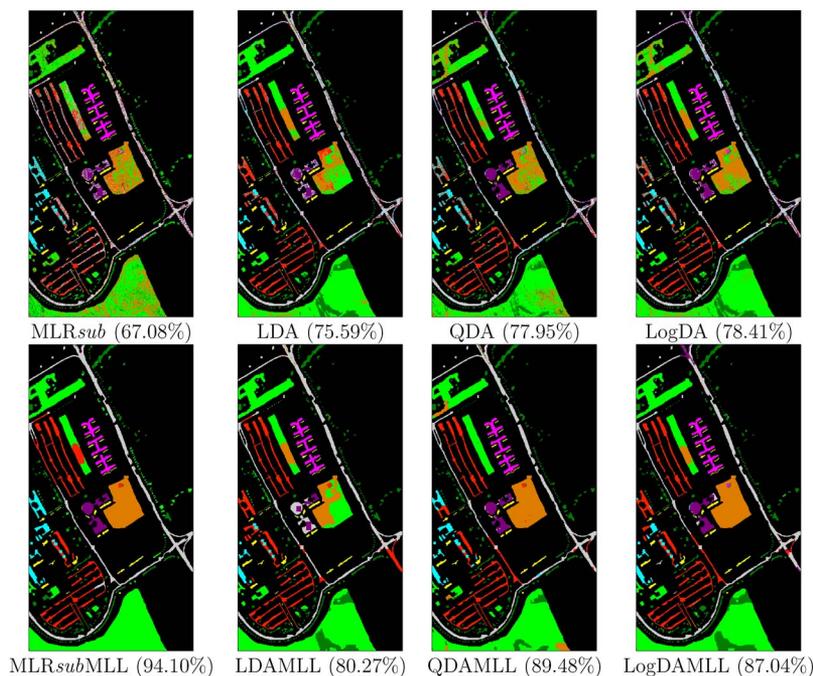
Fig. 11. Classification/segmentation maps obtained by the different tested methods for the ROSIS University of Pavia scene (overall accuracies are reported in parentheses).

TABLE III
OVERALL AND AVERAGE CLASS ACCURACIES (IN PERCENT) AND κ STATISTIC OBTAINED AFTER COMPARING THE PROPOSED MLR*sub*MLL ALGORITHM WITH OTHER SPATIAL–SPECTRAL CLASSIFIERS PRESENTED IN [50] FOR THE ROSIS UNIVERSITY OF PAVIA DATA SET

Classifier	ECHO♯	SVMMRF+MV♮	SVMMRF-NE♭	SVMMRF-E♭	LORSAL-MLL§	MLR*sub*MLL
OA	87.58	91.08	86.89	87.63	85.57	94.10
AA	92.16	94.76	92.12	93.41	92.54	93.45
kappa	83.90	88.30	83.14	84.07	81.80	92.24

♯ ECHO denotes the well-known extraction and classification of homogeneous objects algorithm in [1];
♮ SVMMSF+MV denotes an SVM-based approach based on the construction of a minimum spanning forest followed by majority voting within connected regions [50];
♭ SVMMRF-NE and SVMMRF-E denote two different SVM- and MRF-based spectral-spatial methods using different spatial energy functions [50].
§ LORSAL-MLL denotes our previous work presented in [22].
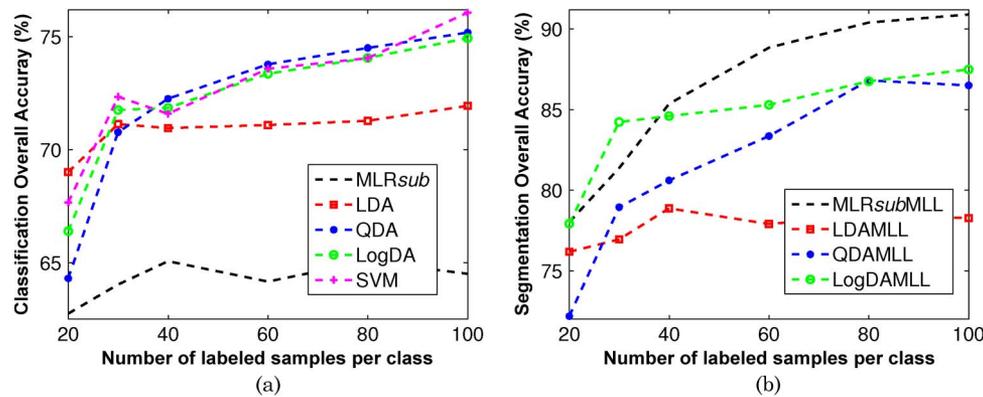
Fig. 12. OA results as a function of the number of labeled samples per class for the University of Pavia data set. (a) Classification results. (b) Segmentation results.

REFERENCES

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

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

[3] S. Serpico and G. Moser, "Extraction of spectral channels from hyperspectral images for classification purposes," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 2, pp. 484–495, Feb. 2007.

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

[5] B. Scholkopf and A. Smola, *Learning With Kernels-Support Vector Machines, Regularization, Optimization and Beyond*. Cambridge, MA: MIT Press, 2002.

[6] 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.

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

[8] 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.

[9] A. Plaza, J. Benediktsson, J. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. S110–S122, Sep. 2009.

[10] S. Watanabe, P. Lambert, C. Kulikowski, J. Buxton, and R. Walker, "Evaluation and selection of variables in pattern recognition," in *Computer and Information Sciences II*. New York: Academic, 1967.

[11] E. Oja, *Subspace Methods of Pattern Recognition*. Lechworth, U.K.: Res. Stud. Press Ltd., 1983.

[12] J. Galambos and I. Simonelli, *Bonferroni-Type Inequalities With Applications*. New York: Springer-Verlag, 1996.

[13] V. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.

[14] A. Ng and M. Jordan, "On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes," in *Proc. 16th Annu. Conf. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, 2002, pp. 841–848.

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

[16] H. Bagan, Y. Yasuoka, T. Endo, X. Wang, and Z. Feng, "Classification of airborne hyperspectral data based on the average learning subspace method," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 368–372, Jul. 2008.

[17] J.-M. Yang, B.-C. Kuo, P.-T. Yu, and C.-H. Chuang, "A dynamic subspace method for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 7, pp. 2840–2853, Jul. 2010.

[18] J. Borges, J. Bioucas-Dias, and A. Marçal, "Evaluation of Bayesian hyperspectral imaging segmentation with a discriminative class learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Barcelona, Spain, 2007, pp. 3810–3813.

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

[20] G. Licciardi, F. Pacifici, D. Tuia, S. Prasad, T. West, F. Giacco, C. Thiel, J. Inglada, E. Christophe, J. Chanussot, and P. Gamba, "Decision fusion for the classification of hyperspectral data: Outcome of the 2008 GRS-S data fusion contest," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3857–3865, Nov. 2009.

[21] Y. Tarabalka, J. Chanussot, and J. Benediktsson, "Segmentation and classification of hyperspectral images using watershed transformation," *Pattern Recognit.*, vol. 43, no. 7, pp. 2367–2379, Jul. 2010.

[22] J. Li, J. Bioucas-Dias, and A. Plaza, "Hyperspectral image segmentation using a new Bayesian approach with active learning," *IEEE Trans. Geosci. Remote Sens.*, 2011, DOI: 10.1109/TGRS.2011.2128330.

[23] R. Fjortoft, Y. Delignon, W. Pieczynski, M. Sigelle, and F. Tupin, "Unsupervised classification of radar images using hidden Markov chains and hidden Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 3, pp. 675–686, Mar. 2003.

[24] Y. Zhao, L. Zhang, P. Li, and B. Huang, "Classification of high spatial resolution imagery using improved Gaussian Markov random-field-based texture features," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1458–1468, May 2007.

[25] Y. Boykov, O. Veksler, and R. Zabih, "Efficient approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 12, pp. 1222–1239, Nov. 2001.

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

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

[28] K. Guilfoyle, M. Althouse, and C.-I. Chang, "A quantitative and comparative analysis of linear and nonlinear spectral mixture models using radial basis function neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 10, pp. 2314–2318, Oct. 2001.

[29] C. Borel and S. Gerslt, "Nonlinear spectral mixing models for vegetative and soil surfaces," *Remote Sens. Environ.*, vol. 47, no. 3, pp. 403–416, 1994.

[30] 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.

[31] D. Hunter and K. Lange, "A tutorial on MM algorithms," *Amer. Statistician*, vol. 58, no. 1, pp. 30–37, 2004.

[32] M. Figueiredo, J. Bioucas-Dias, and R. Nowak, "Majorization–minimization algorithms for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2980–2991, Dec. 2007.

[33] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984.

[34] S. Li, *Markov Random Field Modeling in Image Analysis*. New York: Springer-Verlag, 2001.

[35] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. R. Stat. Soc. B*, vol. 36, no. 2, pp. 192–236, 1974.

[36] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, Sep. 2004.

[37] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 147–159, Feb. 2004, San Mateo, CA: Morgan Kaufmann.

[38] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Understanding belief propagation and its generalizations," in *Exploring Artificial Intelligence in the New Millennium*, vol. 8. San Mateo, CA: Morgan Kaufmann, 2003, pp. 236–239.

[39] J. Yedidia, W. Freeman, and Y. Weiss, "Constructing free energy approximations and generalized belief propagation algorithms," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2282–2312, Jul. 2005.

[40] V. Kolmogorov, "Convergent tree-reweighted message passing for energy minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1568–1583, Oct. 2006.

[41] S. Bagon, *Matlab Wrapper for Graph Cut*, Dec. 2006. [Online]. Available: http://www.wisdom.weizmann.ac.il/~bagon

[42] R. Duda, P. Hart, and D. Stork, *Pattern Classification*. New York: Wiley-Interscience, 2000.

[43] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. Int. Joint Conf. Artif. Intell.*, 1995, pp. 1137–1143.

[44] R. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, vol. 7, pp. 179–188, 1936.

[45] B. Ripley, *Exploring artificial intelligence in the new millennium*. Cambridge, U.K.: Cambridge Univ. Press, 1966.

[46] J. Poulsen and A. French, *Discriminant Function Analysis*. [Online]. Available: http://userwww.sfsu.edu/~efc/classes/biol710/discrim/discrim.pdf

[47] T. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of hyperspectral images with regularized linear discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 862–873, Mar. 2009.

[48] Q. Du and N. Younan, "On the performance improvement for linear discriminant analysis-based hyperspectral image classification," in *Proc. IAPR Workshop Pattern Recognit. Remote Sens.*, 2008, pp. 1–4.

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

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

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

José M. Bioucas-Dias (S'87–M'95) received the E.E., M.Sc., Ph.D., and "Agregado" degrees in electrical and computer engineering from the Technical University of Lisbon, Lisbon, Portugal, in 1985, 1991, 1995, and 2007, respectively.

Since 1995, he has been with the Department of Electrical and Computer Engineering, Instituto Superior Tecnico, Technical University of Lisbon. He is also a Senior Researcher with the Pattern and Image Analysis Group, Instituto de Telecomunicações, which is a private nonprofit research institution. His research interests include signal and image processing, pattern recognition, optimization, and remote sensing. He was and is involved in several national and international research projects and networks, including the Marie Curie Actions "Hyperspectral Imaging Network (Hyper-I-Net)" and the "European Doctoral Program in Signal Processing (SIGNAL)."

Dr. Bioucas-Dias is an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING, and he was an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS. He is a Guest Editor of two IEEE special issues (IEEE TGRS and IEEE JSTARS). He has been a member of program/technical committees of several international conferences. He was the General Cochair of the 3rd IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS 2011).

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 50 journal citation report papers, around 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. He has coedited a book on high-performance computing in remote sensing and guest edited four 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*.

Dr. Plaza 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 Europe Conference on High Performance Computing in Remote Sensing since 2011. He 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 for the IEEE Geoscience and Remote Sensing Society.

Jun Li received the B.S. degree in geographic information systems from Hunan Normal University, Hunan, China, in 2004 and the M.E. degree in remote sensing from Peking University, Beijing, China, in 2007.

From 2007 to 2010, she was a Marie Curie Research Fellow with the Departamento de Engenharia Electrotécnica e de Computadores, Instituto de Tele-comunicações, Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisboa, Portugal, in the framework of the European Doctorate for Signal Processing (SIGNAL) under the joint supervision of Prof. José M. Bioucas-Dias and Prof. Antonio Plaza. Currently, she is with the Hyperspectral Computing Laboratory (HyperComp) research group coordinated by Prof. Antonio Plaza at the Department of Technology of Computers and Communications, University of Extremadura, Cáceres, Spain. Her research interests include hyperspectral image classification and segmentation, spectral unmixing, signal processing, and remote sensing. She has been a Reviewer of several journals, including *Optical Engineering* and *Inverse Problems and Imaging*.

Ms. Li is a Reviewer for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.