Spectral–Spatial Classification of Hyperspectral Data Using Local and Global Probabilities for Mixed Pixel Characterization

Mahdi Khodadadzadeh, *Student Member, IEEE*, Jun Li, *Member, IEEE*, Antonio Plaza, *Senior Member, IEEE*, Hassan Ghassemian, *Senior Member, IEEE*, José M. Bioucas-Dias, *Member, IEEE*, and Xia Li

***Abstract*—Remotely sensed hyperspectral image classification is a very challenging task. This is due to many different aspects, such as the presence of mixed pixels in the data or the limited information available *a priori*. This has fostered the need to develop techniques able to exploit the rich spatial and spectral information present in the scenes while, at the same time, dealing with mixed pixels and limited training samples. In this paper, we present a new spectral–spatial classifier for hyperspectral data that specifically addresses the issue of mixed pixel characterization. In our presented approach, the spectral information is characterized both locally and globally, which represents an innovation with regard to previous approaches for probabilistic classification of hyperspectral data. Specifically, we use a subspace-based multinomial logistic regression method for learning the posterior probabilities and a pixel-based probabilistic support vector machine classifier as an indicator to locally determine the number of mixed components that participate in each pixel. The information provided by local and global probabilities is then fused and interpreted in order to characterize mixed pixels. Finally, spatial information is characterized by including a Markov random field (MRF) regularizer. Our experimental results, conducted using both synthetic and real hyperspectral images, indicate that the proposed classifier leads to state-of-the-art performance when compared with other approaches, particularly in scenarios in which very limited training samples are available.**

***Index Terms*—Hyperspectral imaging, Markov random field (MRF), multiple classifiers, spectral–spatial classification, subspace multinomial logistic regression (MLR*sub*), support vector machine (SVM).**

I. Introduction

HYPERSPECTRAL imaging instruments are now able to collect hundreds of images, corresponding to different wavelength channels, for the same area on the surface of the Earth [1]. Hyperspectral image classification has been a very active area of research in recent years [2]. Given a set of observations (i.e., pixel vectors in a hyperspectral image), the goal of classification is to assign a unique label to each pixel vector so that it is well defined by a given class [3]. The availability of hyperspectral data with high spectral resolution has been quite important for many applications, such as crop mapping, environmental monitoring, and object identification for defense purposes [4].

Several techniques have been used to perform supervised classification of hyperspectral data. Classic techniques include maximum likelihood (ML) [2], [3], [5], nearest neighbor classifiers [6], or neural networks [7]–[9], among many others [4]. The quality of these pixelwise classification methods is strongly related to the quality and number of training samples. In order to effectively learn the parameters of the classifier, a sufficient number of training samples are required. However, training samples are difficult and expensive to collect in practice [10]. This issue is quite problematic in hyperspectral analysis, in which there is often an unbalance between the high dimensionality of the data and the limited number of training samples available in practice, known as the Hughes effect [2].

In this context, kernel methods such as the support vector machine (SVM) have been widely used in hyperspectral imaging to deal effectively with the Hughes phenomenon by addressing large input spaces and producing sparse solutions [11]–[14]. Recently, the multinomial logistic regression (MLR) [15] has been shown to provide an alternative approach to deal with ill-posed problems. This approach has been explored in hyperspectral imaging as a technique able to model the posterior class distributions in a Bayesian framework, thus supplying (in addition to the boundaries between the classes) a degree of plausibility for such classes [15]. A main difference between the MLR and other classifiers such as the probabilistic SVM is the fact that the former learns the posterior probabilities for the whole image. As a result, these classifiers exploit the probabilistic information in a different (possibly complementary) fashion, although this issue has never been explored in the literature in the past. Recently, the advantages of probabilistic SVM as a soft classification technique in discriminating between pure and mixed pixels, and automatically selecting endmember subsets were, respectively, investigated in [16] and [17]. These techniques pay particular attention to characterizing the number of mixtures participating in each pixel.

A subspace-based version of the MLR classifier, called MLR*sub* [18], has also been recently developed. This method

Manuscript received August 1, 2013; revised November 29, 2013; accepted December 6, 2013. Date of publication March 5, 2014; date of current version May 22, 2014. This work was supported by the National Basic Research Program of China (973 Program, Grant 2011CB707103). *(Corresponding author: J. Li.)*

M. Khodadadzadeh and A. Plaza are with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, 10003 Cáceres, Spain.

J. Li and X. Li are with the School of Geography and Planning and Guangdong Key Laboratory for Urbanization and Geo-Simulation, Sun Yat-Sen University, Guangzhou 510275, China (e-mail: jun@lx.it.pt).

H. Ghassemian is with the Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Tehran 14155-4843, Iran.

J. M. Bioucas-Dias is with the Telecommunications Institute, Instituto Superior Técnico, 1049-001 Lisbon, Portugal.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2013.2296031

relies on the basic assumption that the samples within each class can approximately lie in a lower dimensional subspace and uses subspace projection methods to find this subspace. Since hyperspectral data are likely to be noisy and dominated by mixed pixels, the MLR*sub* has been shown to provide good performance (particularly, in the case of limited training samples) as normally classes live in a much lower space in comparison with the original data dimensionality.

Another strategy to deal with the limited number of training samples available in practice has been to efficiently exploit labeled information by using multiple classifier systems or classifier *ensembles* [19]–[22]. This approach has been proved successful in different hyperspectral image classification applications [23]–[26].

Finally, a well-known trend in order to alleviate the problem of insufficient number of training samples is to integrate the spatial–contextual information in the analysis. Many examples of spectral–spatial classifiers can be found in the hyperspectral imaging literature [4], [27]–[33]. In particular, approaches based on Markov random fields (MRFs) have been quite successful in hyperspectral imaging [15], [18], [34]–[37]. In particular, [36] successfully combined a probabilistic SVM with an MRF regularizer for the classification of hyperspectral images. All of these methods exploit, in a way or another, the continuity (in probability sense) of neighboring labels. In other words, these methods exploit the likely fact that, in a hyperspectral image, two neighboring pixels may have the same label.

In this paper, we propose a new spectral–spatial classifier in which the spectral information is characterized both locally and globally. Specifically, we use the MLR*sub* method to globally and locally learn the posterior probabilities for each pixel, where addressing the local probability is one of the main innovative contributions of this work. For local probability learning, we determine the number of mixed components that participate in each pixel. For this purpose, we use a probabilistic SVM as an indicator to determine the number of mixed components. Finally, the spatial information is then characterized by exploiting an MRF regularizer.

When compared to the probabilistic SVM, the presented classifier considers mixtures in the model. This is very important since hyperspectral images are often dominated by mixed pixels. When compared to the MLR*sub*, which already addresses the presence of mixed pixels, the proposed classifier constrains the number of mixed components, thus improving its characterization since mixed pixels in hyperspectral images normally comprise only a few mixing components [38]. As a result, the presented approach provides two important contributions with regard to existing spectral–spatial approaches. The first one is the consideration of probabilistic information at both local and global levels. The second one is the characterization of the number of mixtures participating in each pixel, which is quite important since mixed pixels often dominate hyperspectral data.

The presented approach also observes two of the most pressing needs of current hyperspectral classifiers: the possibility to use very limited training sets (compensated by the multiple classifier flavor of our approach) and the need to integrate spatial information in the assessment (addressed by the inclusion of an MRF regularizer in the formulation). The resulting method, called SVM-MLR*sub*-MRF, achieves very good classification results which are competitive or superior to those provided by many other state-of-the-art supervised classifiers for hyperspectral image analysis.

The remainder of this paper is organized as follows. Section II describes the different strategies used to implement the proposed spectral–spatial classifier. Section III describes the proposed approach. An important observation is that the presented approach should not be simply understood as a combination of existing approaches. Specifically, each of the processing algorithms described in Section III corresponds to one out of many possible choices, selected based on their availability and also on the possibility to draw comparisons with other established techniques for spectral–spatial classification. However, it should be noted that other strategies for addressing local versus global information for mixed pixel characterization and spatial regularization could be used. In this regard, our selection should be strictly understood as a vehicle to demonstrate a new framework for classification of hyperspectral data and not merely as a combination of processing blocks. To the best of our knowledge, the presented framework addresses for the first time in the literature the aforementioned aspects in synergistic fashion. Section IV presents extensive experiments using both simulated and real hyperspectral data designed in order to validate the method and provide comparisons with other state-of-the-art classifiers. Section V concludes with some remarks and hints at plausible future research lines.

II. Main Components of the Proposed Method

In this section, we describe the different components that have been used in the development of the proposed method. First, we use probabilistic pixelwise classification methods to learn the posterior probability distributions from the spectral information. Here, we use two strategies to characterize spectral information: probabilistic SVM and MLR*sub*. Then, we use contextual information by means of an MRF regularization scheme to refine the classification results. Fig. 1 shows the relationship between the methods considered in this study. As it can be observed from Fig. 1, estimating class conditional probability distributions is an intrinsic issue for the subsequent MRF-based classification. In the following, we outline the different strategies used to characterize spectral and spatial information, respectively, in the presented approach.

A. Characterization of Spectral Information

Let $\mathbf{x} \equiv \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ be the input hyperspectral image, where $\mathbf{x}_i = [\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{id}]^T$ denotes a spectral vector associated with an image pixel $i \in S$, $S = \{1, 2, \ldots, n\}$ is the set of integers indexing the n pixels of $\mathbf{x}$, and d is the number of spectral bands. Let $\mathbf{y} \equiv (\mathbf{y}_1, \ldots, \mathbf{y}_n)$ denote an image of class labels $\mathbf{y}_i \equiv [y_{i1}, y_{i2}, \ldots, y_{ic} \ldots, y_{ik}]^T$, where k is the number of classes, $y_{ic} = \{0, 1\}$, for $c = 1, \ldots, k$, and $\sum_c y_{ic} = 1$. Furthermore, let $\mathcal{D}_l \equiv \{(\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_l, \mathbf{y}_l)\}$ be the labeled training set, with l being the number of samples in $\mathcal{D}_l$. With the aforementioned definitions in mind, probabilistic pixelwise

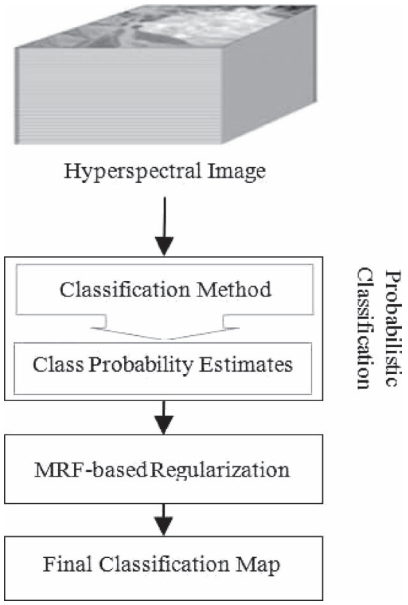

Fig. 1. Relationship between the different components of the presented approach for spectral–spatial classification of hyperspectral data.

classification intends to obtain, for a given pixel $\mathbf{x}_i$, the class label vector $\mathbf{y}_i$. This vector can be obtained by computing the posterior probability $p(y_{ic} = 1|\mathbf{x}_i, \mathcal{D}_l)$ as follows:

$$y_{ic} = \begin{cases} 1, & \text{if } p(y_{ic} = 1|\mathbf{x}_i, \mathcal{D}_l) > p(y_{ic_t} = 1|\mathbf{x}_i, \mathcal{D}_l) \\ & \quad \forall\, c_t \neq c \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Various probabilistic classification techniques have been used to process hyperspectral data [3]. In this paper, we use the probabilistic SVM [39] and the MLR*sub* classifiers [18] for probability estimation. SVMs and MLR*sub* rely, respectively, on discriminant functions and posterior class distributions which have shown good performance in hyperspectral data classification, particularly in scenarios dominated by small training samples. In the following, we describe these probabilistic classifiers in more details.

1) Probabilistic SVM Algorithm: The SVM classifier is typically defined as follows [39]:

$$f(\mathbf{x}_j) = \sum_i \alpha_i y_i \Phi(\mathbf{x}_i, \mathbf{x}_j) + b \tag{2}$$

where $\mathbf{x}_j \in \mathbf{x}$, $\mathbf{x}_i \in \mathcal{D}_l$, b is the bias, and $\{\alpha_i\}_{i=1}^{l}$ represents Lagrange multipliers which are determined by the parameter C (that controls the amount of penalty during the SVM optimization). Here, $\Phi(\mathbf{x}_i, \mathbf{x}_j)$ is a function of the inputs, which can be linear or nonlinear. In SVM classification, kernel methods have shown great advantage in comparison with linear methods [14]. In this paper, we use a Gaussian radial basis function kernel $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2)$, whose width is controlled by parameter γ. Although the original SVM does not provide class probability estimates, different techniques can be used to obtain class probability estimates based on combining all pairwise comparisons [39]. In this paper, one of the probabilistic SVM methods [41] included in the popular LIBSVM library [42] is used.

2) MLRsub Algorithm: MLR-based techniques are able to model the posterior class distributions in a Bayesian framework. In these approaches, the densities $p(\mathbf{y}_i|\mathbf{x}_i)$ are modeled with the MLR, which corresponds to discriminative model of the discriminative–generative pair for $p(\mathbf{x}_i|\mathbf{y}_i)$ Gaussian and $p(\mathbf{y}_i)$ multinomial. The MLR model is formally given by [43]

$$p(y_{ic} = 1|\mathbf{x}_i, \boldsymbol{\omega}) = \frac{\exp\left(\boldsymbol{\omega}^{(c)}\mathbf{h}^{(c)}(\mathbf{x}_i)\right)}{\sum_{c=1}^{k}\exp\left(\boldsymbol{\omega}^{(c)}\mathbf{h}(\mathbf{x}_i)\right)} \tag{3}$$

where $\mathbf{h}^{(c)}(\mathbf{x}_i) \equiv [\mathbf{h}_1^{(c)}(\mathbf{x}_i), \ldots, \mathbf{h}_l^{(c)}(\mathbf{x}_i)]^T$ is a vector of l fixed functions of the input data, often termed as features; $\boldsymbol{\omega}^{(c)}$ is the set of logistic regressors for class c, and $\boldsymbol{\omega} \equiv [\boldsymbol{\omega}^{(1)^T}, \ldots, \boldsymbol{\omega}^{(c-1)^T}]^T$. Recently, Li *et al.* [18] have proposed to combine MLR with a subspace projection method called MLR*sub* to cope with two main issues: the presence of mixed pixels in hyperspectral data and the availability of limited training samples. 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 [18]. In this formulation, the input function is class dependent and is given by

$$\mathbf{h}^{(c)}(\mathbf{x}_i) = \left[\|\mathbf{x}_i\|^2, \left\|\mathbf{x}_i^{\mathrm{T}}\mathbf{U}^{(c)}\right\|^2\right]^{\mathrm{T}} \tag{4}$$

where $\mathbf{U}^{(c)} = \{\mathbf{u}_1^{(c)}, \ldots, \mathbf{u}_{r^{(k)}}^{(c)}\}$ is a set of $r^{(k)}$-dimensional orthonormal basis vectors for the subspace associated with class c ($r^{(c)} \ll d$).

B. Characterization of Spatial Information

In this section, we describe the mechanism used to include spatial–contextual information in the presented method. For this purpose, we use MRF, which is a widely used contextual model and a classical probabilistic method to model spatial correlation of pixel neighbors. This approach has been successfully applied in the context of remote sensing problems [35], [37], [41]. In the MRF framework, the MAP decision rule is typically formulated as the minimization of a suitable energy function [34]. Normally, the MRF-based approach can be implemented in two steps in hyperspectral image analysis. First, a probabilistic pixelwise classification method (such as those described in the previous section) is applied to learn the posterior probability distributions from the spectral information. Second, contextual information is included by means of an MRF regularization to refine the classification, as already outlined in Fig. 1.

According to the MAP-MRF framework, a pixel belonging to a class c is very likely to have neighboring pixels belonging to the same class. By using the Hammersley–Clifford theorem [45], we can write the MAP estimate of $\mathbf{y}$ as follows:

$$\widehat{\mathbf{y}} = \arg\min_{\mathbf{y}} \left(\sum_{i \in \mathcal{S}} -\log p(\mathbf{y}_i|\mathbf{x}_i) - \mu \sum_{i \sim j} \delta(\mathbf{y}_i - \mathbf{y}_j) \right) \tag{5}$$

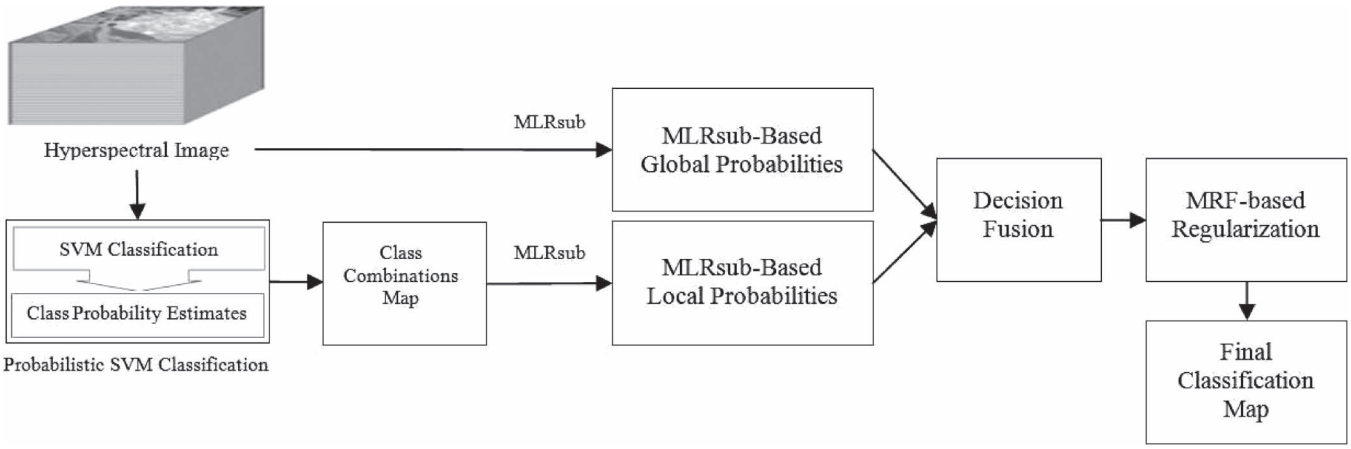

Fig. 2. Flowchart of the proposed SVM-MLR*sub*-MRF method.

where the term $p(\mathbf{y}_i|\mathbf{x}_i)$ is the spectral energy function from the observed data, which needs to be estimated by probabilistic methods. In this paper, we use the probabilistic SVM and MLR*sub* to learn the probabilities. Parameter μ is tunable and controls the degree of smoothness, and $\delta(y)$ is the unit impulse function, where $\delta(0) = 1$ and $\delta(y) = 0$ for $y \neq 0$. Notice that the pairwise terms, $\delta(\mathbf{y}_i - \mathbf{y}_j)$, attach higher probability to equal neighboring labels than the other way around. Minimization of expression (5) is a combinatorial optimization problem involving unary and pairwise interaction terms. A good approximation can be obtained by mapping the problem into the computation of a series of min-cuts on a suitable graphs [45]. This aspect has been thoroughly explored in the context of hyperspectral image classification in previous contributions [15].

III. Proposed Approach

In this section, we present the proposed spectral–spatial classification approach called SVM-MLR*sub*-MRF. The full methodology is summarized by a detailed flowchart in Fig. 2. As shown in Fig. 2, the proposed approach mainly contains four steps: 1) generation of the class combination map; 2) calculation of the local and global probabilities; 3) decision fusion; and 4) MRF regularization. In the following, we present the details of each individual steps.

A. Generation of the Class Combination Map

The class combination map is generated from the probabilistic SVM classification results. Notice that the probabilistic SVM is only used as an indicator to determine the number of mixtures appearing in each pixel and does not contribute to the probability learning. For this purpose, a subset of the M most reliable class labels (mixed components) is chosen for each pixel as the possible class combinations for that pixel, and $M \leq k$. In case $M = k$, the local learning is equivalent to the global learning. It is also important to emphasize that, although in this work we use the probabilistic SVM for pixelwise classification due to its proved effectiveness in hyperspectral classification [14], other probabilistic classifiers could also be used as far as they are well suited to hyperspectral analysis. Furthermore, as a classifier, the probabilistic SVM has different characteristics

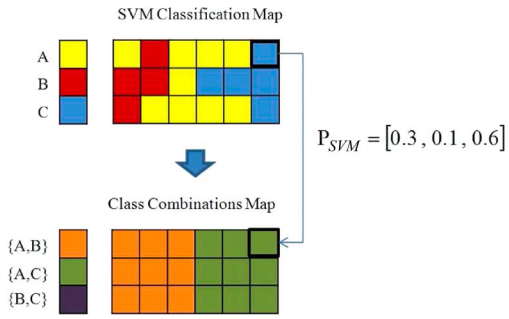

Fig. 3. Example of the generation of a class combination map using threshold $M = 2$.

in comparison with MLR*sub*, thus allowing for the possibility to use both classifiers in combined fashion in order to remove irrelevant class labels and to improve the efficiency of the class-dependent subspace projection step in the MLR*sub* method, which will be described in the following section.

For illustrative purposes, Fig. 3 shows an example of how to generate a class combination map using the aforementioned strategy for a three-class problem, where the classes are denoted as $\{A, B, C\}$ and the number of mixed components is set to $M = 2$. Using the probabilistic SVM, for each pixel, we obtain a vector of three probabilities with respect to classes A, B, and C. As shown in Fig. 3, for the pixel at the top-right corner of the image, we assume that the probabilities are 0.3, 0.1, and 0.6 (for classes A, B, and C, respectively). Under these assumptions, the pixel would be assigned to the subset $\{A, C\}$ (from all possible combinations of the three classes). Notice that, in this example, there is no pixel assigned to the class combination $\{B, C\}$. Finally, it should be noted that the number of class combinations is given by $C(k, M)$, where, in this example, it is $C(3, 2) = 3$.

B. Calculation of the Local and Global Probabilities

In this section, we describe the procedure used to calculate the local and global probabilities. Here, we use the MLR*sub* algorithm to learn the posterior probability distributions locally for the M classes selected in the previous step and globally for all classes. Let $\mathbf{p}_g$ and $\mathbf{p}_l$ denote the global and local posterior probabilities, respectively. For example, if we take the pixel

used as an example in the previous section (i.e., the one located at the top-right corner in Fig. 3), in this case, compute the global and local probabilities as follows:

$$\mathbf{p}_g = \{p(y_{ic} = 1|\mathbf{x}_i, \boldsymbol{\omega}_g), c = A, B, C\} \quad (6)$$

$$\mathbf{p}_l = \{p(y_{iA} = 1|\mathbf{x}_i, \boldsymbol{\omega}_l), 0, p(y_{iC} = 1|\mathbf{x}_i, \boldsymbol{\omega}_l)\} \quad (7)$$

where (6) is the global learning step and $\boldsymbol{\omega}_g$ represents the corresponding logistic regressors. On the other hand, (7) is the local learning step, and $\boldsymbol{\omega}_l$ represents the associated regressors. The global probability $\mathbf{p}_g$ in (6) is learned from the original data by the MLR*sub* algorithm [18]. Here, we exploit the good capabilities of MLR*sub* when dealing with mixtures. At this point, it is important to emphasize that we selected the MLR*sub* because, in real images, it is very likely that an observed pixel is a mixture of several components/classes. However, it is unlikely that the pixel is mixed by many components/classes [38]. Based on this observation, (7) uses MLR*sub* to locally learn the class posterior probabilities from the class combination map generated by the probabilistic SVM. Notice that, in the local estimation, only M classes are considered for each pixel, and we remove the remaining ones, i.e., their probabilities are set to 0. For instance, in (7), $p(y_{ic} = 1|\mathbf{x}_i, \boldsymbol{\omega}_l) = 0$, which means that c is a removed class. In this way, by setting the probability of the irrelevant classes to zero, we remove those irrelevant classes from the combination set such that we eliminate the influence of the less relevant classes (or noise) in the local area. Therefore, considering the mixtures at a local (pixel) level is very important due to the following reasons. First, by eliminating the less relevant classes, the proposed approach locally eliminates noise, which greatly improves the separability of the features from noise, thus improving the performance of the MLR*sub* algorithm. Second, the class-dependent subspace reduces its dimensionality such that less training samples are required for learning.

C. Decision Fusion

In this step, we combine the local and global probability distributions learned by the MLR*sub* algorithm to produce the final probabilities. For this purpose, we use the consensus theory [46] which follows a linear opinion pool [47]:

$$p(y_{ic} = 1|\mathbf{x}_i) = \sum_{j=1}^{N} \lambda_j p_j(y_{ic} = 1|\mathbf{x}_i) \quad (8)$$

where N is the number of data sources, $p_j(y_{ic} = 1|\mathbf{x}_i)$ is a source-specific posterior probability associated with data source j, and λ_j's are source-specific weights which control the relative influence of the data sources, where $0 \leq \lambda_j \leq 1$ and $\sum_{j=1}^{N} \lambda_j = 1$. In this paper, we consider two data sources: global and local probability distributions. We simply combine these two data sources to compute the final class probability estimates as

$$p(y_{ic}=1|\mathbf{x}_i)=\lambda p_g(y_{ic}=1|\mathbf{x}_i, \boldsymbol{\omega}_g)+(1-\lambda)p_l(y_{ic}=1|\mathbf{x}_i, \boldsymbol{\omega}_l) \quad (9)$$

where λ is a tunable parameter which controls the weights between the global and local probabilities and $0 \leq \lambda \leq 1$. It should be noted that, if $\lambda = 1$, only the global information is considered and the method remains as the original MLR*sub*. If $\lambda = 0$, only the local information is used. In our experimental results section, we will analyze the impact of parameter λ and discuss the relevance of global information and local information in the obtained classification results.

D. MRF-Based Spatial Regularization

The last step of our proposed method consists of including the spatial–contextual information. As shown by Fig. 2, this stage is applied on the output of the decision fusion step. Although many strategies can be used for this purpose, we follow a commonly used strategy which relies on an MAP-MRF framework [18], as described in Section II-B. The strategy adopted in this work is similar to the one adopted by the MLR*sub*-MRF [15], which will be used for comparative purposes in the following section. Our strategy is also similar to the one used by the SVM-MRF in [35] but emphasizes that, in our comparisons, we do not exactly use the algorithm introduced in [35] but an implementation of SVM-MRF developed by ourselves and based on graph-cuts [45].

IV. Experimental Results

In this section, we use both synthetic and real hyperspectral data sets to evaluate the performance of the proposed SVM-MLR*sub*-MRF classification algorithm in different analysis scenarios. The main objective of the experimental validation with synthetic hyperspectral image is the assessment and characterization of the algorithm in a fully controlled environment, whereas the main objective of the experimental validation with real data sets is to compare the performance of the proposed method with other state-of-the-art methods in the literature. The remainder of this section is organized as follows. First, we describe the hyperspectral data sets (synthetic and real) used in experiments. Then, we describe the experimental setting. Next, we describe several experiments intended to address several important aspects of the presented method, such as the impact of parameters λ and M, a comparison with other standard methods, and an evaluation of the performance of the method in the presence of training sets with different numbers of samples.

A. Hyperspectral Data Sets

1) Synthetic Data: A synthetic image has been generated with a size of $n = 80 \times 120$, and the class distribution is displayed in Fig. 4(a). The synthetic scene comprises eight classes which contain linear mixtures of a set of spectral signatures randomly selected from a digital spectral library compiled by the U.S. Geological Survey (USGS) and available online.[1] The USGS library contains spectral plots for nearly 500 materials (mostly minerals) in the 400–2500-nm spectral range, where the bands have been convolved to the number of bands available for the Airborne Visible/Infrared Imaging

[1] http://speclab.cr.usgs.gov/spectral-lib.html

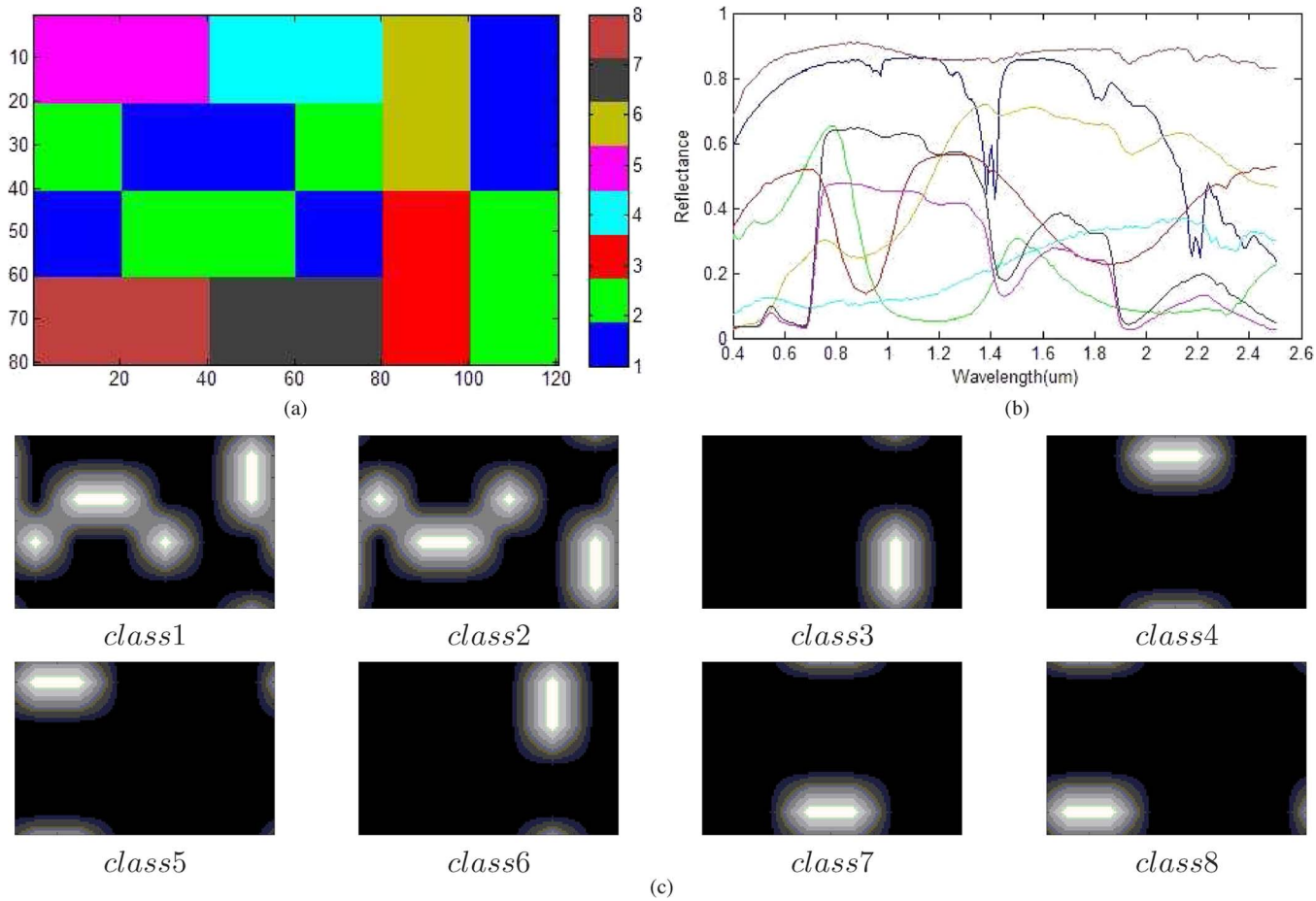

Fig. 4. (a) Classes in a synthetic scene with $n = 80 \times 120$. (b) Spectral signatures of randomly selected materials from the USGS digital library used in the simulation. (c) Fractional abundance distributions considered for generating mixed pixels using a Gaussian filter of size $k = 25$ and standard deviation $\sigma = 30$.

Spectrometer (AVIRIS) [48] that comprises 224 spectral bands. Fig. 4(b) shows the spectral signatures of eight randomly selected mineral signatures allocated to the main classes displayed in Fig. 4(a).

In order to simulate mixed pixels using linear mixtures, the eight class abundance maps are filtered by a $k \times k$ symmetric Gaussian low-pass filter with a fixed standard deviation σ. For illustrative purposes, Fig. 4(c) shows the abundance maps associated to the eight classes of the synthetic scene after applying a Gaussian filter of size $k = 25$ and standard deviation $\sigma = 20$. In each pixel of the scene, the fractional abundances vary from 0% (black color) to 100% (white color) and sum to unity. Using this procedure, signature abundance is not constant over class regions, and the pixels closer to the borders of the regions are more heavily mixed, as expected in real scenarios. Finally, zero-mean Gaussian noise is added to each band of the synthetic hyperspectral image so that the signal-to-noise ratio is equal to 20 dB according to the definition given in [49].

2) Real Data: Two real hyperspectral data sets are used to evaluate the proposed approach. The first one is the well-known AVIRIS Indian Pines scene [see Fig. 5(a)], collected over Northwestern Indiana in June 1992 [2]. The scene is available online[2] and contains 145 $\times$ 145 pixels and 220 spectral bands between 0.4 and 2.5 μm. A total of 20 spectral bands was removed prior to the experiments due to noise and water absorption in those channels. The ground-truth image displayed in Fig. 5(b) contains 10 366 samples and 16 mutually exclusive classes having 20–2468 samples. These data are widely used as a benchmark for testing the accuracy of hyperspectral data classification algorithms, mainly because it constitutes a challenging classification problem due to the presence of mixed pixels in available classes and also because of the unbalanced number of available labeled pixels per class.

We have also used a scene collected by the Reflective Optics Spectrographic Imaging System (ROSIS) for evaluation purposes. These data were acquired over the urban area of the University of Pavia, Pavia, Italy. The flight was operated by the Deutschen Zentrum for Luftund Raumfahrt (DLR; the German Aerospace Agency) in the framework of the HySens project, managed and sponsored by the European Commission. The image size in pixels is 610 $\times$ 340, with very high spatial resolution of 1.3 m/pixel. The number of data channels in the acquired image is 103 (with spectral range from 0.43 to 0.86 μm). Fig. 6(a) shows a false color composite of the image, while Fig. 6(c) shows nine ground-truth classes of interest, which comprise urban features, as well as soil and vegetation features. In the original data set, out of the available ground-truth pixels, 3921 were used for training, and 42 776 samples were used for testing [see Fig. 6(b)].

[2]https://engineering.purdue.edu/~biehl/MultiSpec/

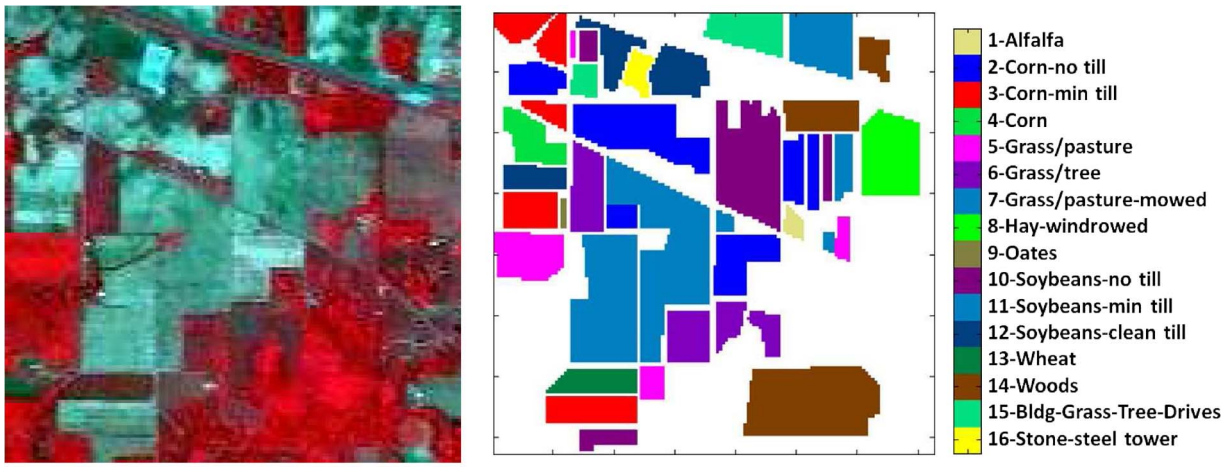

Fig. 5. AVIRIS Indian Pines data set. (a) False color composition. (b) Ground truth as a collection of mutually exclusive classes.

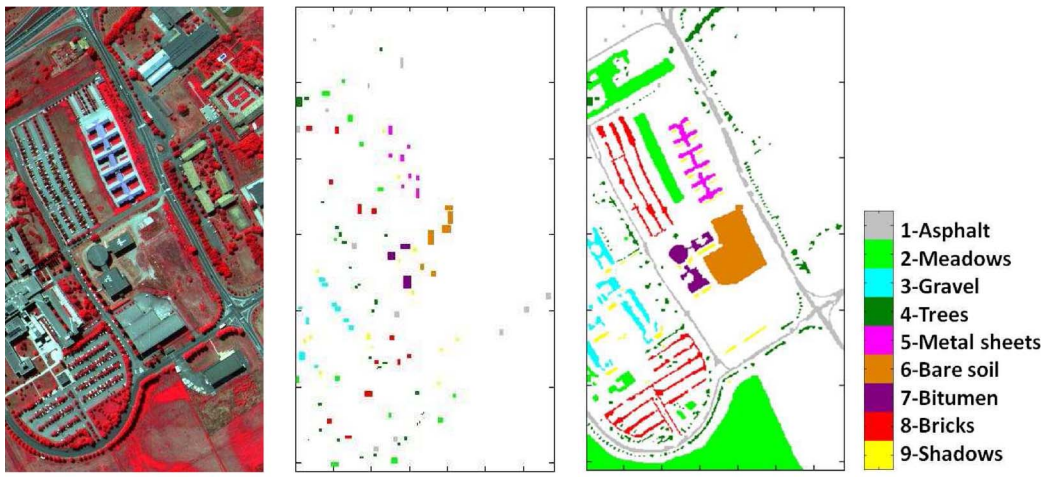

Fig. 6. ROSIS Pavia University data set. (a) False color composition. (b) Training data. (c) Ground truth as a collection of mutually exclusive classes.

B. Experimental Setting

Before describing our results, it is first important to discuss some considerations that define our experimental setting. For the synthetic image experiments, we considered two strategies. In our first strategy, we simulated some pure pixels for each class (using $k = 20$ and $\sigma = 30$ in the synthetic data simulation). In the second strategy, we increased the size of the filter to $k = 25$ (with the same $\sigma = 30$) so that all of the simulated pixels inside a class region were mixed with abundance fractions less than 80%, and the simulated image did not contain any pure pixels. In both cases, training samples were extracted from the purest available classes. All of the results reported in this paper with synthetic data sets were obtained after 50 Monte Carlo runs in which we randomly select 8 different materials and also randomly select different training sets.

Concerning our real data experiments, the experimental setting can be briefly summarized as follows. For the experiments with the AVIRIS Indian Pines data, the training samples were randomly selected from the available ground truth, and the remaining samples are used for validation. For the smallest classes of this data set, if the total number of available labeled samples per class in the ground truth is smaller than the given number of training samples, we take half of the total samples for training. For the ROSIS Pavia University data set, the training sets are composed of subsets of the original training samples, and the remaining test sets are used for validation. All of the results reported in this paper with the two considered real data sets were obtained after 30 Monte Carlo runs.

The classifiers compared in the presented study are the proposed SVM-MLR*sub*-MRF in addition to the standard SVM, MLR*sub*, and SVM-MRF. In all experiments for the MLR*sub* and MLR*sub*-MRF algorithms, we optimized the parameters as indicated in [18]. Concerning the probabilistic SVM classifier, we optimized the related parameters using tenfold cross validation. Finally, for the proposed approach, we use SVM-MLR*sub* and SVM-MLR*sub*-MRF to denote the algorithms with and without the MRF spatial regularizer, respectively.

C. Quantitative and Comparative Assessment

In this section, we conduct an experimental assessment of the presented approach using the simulated and real data sets

TABLE I
OVERALL (OA) AND AVERAGE (AA) CLASSIFICATION ACCURACIES (IN PERCENT; AS A FUNCTION OF PARAMETER λ) OBTAINED BY THE SVM-MLR*sub*-MRF METHOD FOR THE SYNTHETIC AND REAL DATA SETS CONSIDERED IN THE EXPERIMENTS. THE BEST RESULTS ARE OUTLINED IN BOLD TYPEFACE

Data	Accuracies	λ										
		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Synthetic	OA	70.05	70.40	70.63	71.34	72.60	73.61	**73.99**	73.26	72.56	71.48	70.55
(pure training samples)	AA	68.15	68.47	68.74	69.48	71.15	72.84	**73.93**	73.45	73.13	72.19	71.33
Synthetic	OA	75.13	75.66	76.12	76.68	77.82	79.27	**79.40**	78.63	77.36	76.16	75.08
(mixed training samples)	AA	74.54	75.06	75.49	76.13	77.48	79.13	**79.75**	79.13	77.99	76.92	75.85
AVIRIS Indian Pines	OA	84.03	89.39	90.56	91.26	91.83	92.24	**92.38**	92.35	92.29	92.12	91.16
	AA	89.00	92.97	93.93	94.49	94.91	95.08	**95.21**	95.13	95.02	94.81	94.01
Pavia University	OA	90.02	91.78	92.14	92.36	92.48	92.66	**92.70**	92.38	91.72	90.36	87.50
	AA	90.65	91.53	91.62	91.68	91.79	**91.85**	91.78	91.44	90.87	90.18	88.23

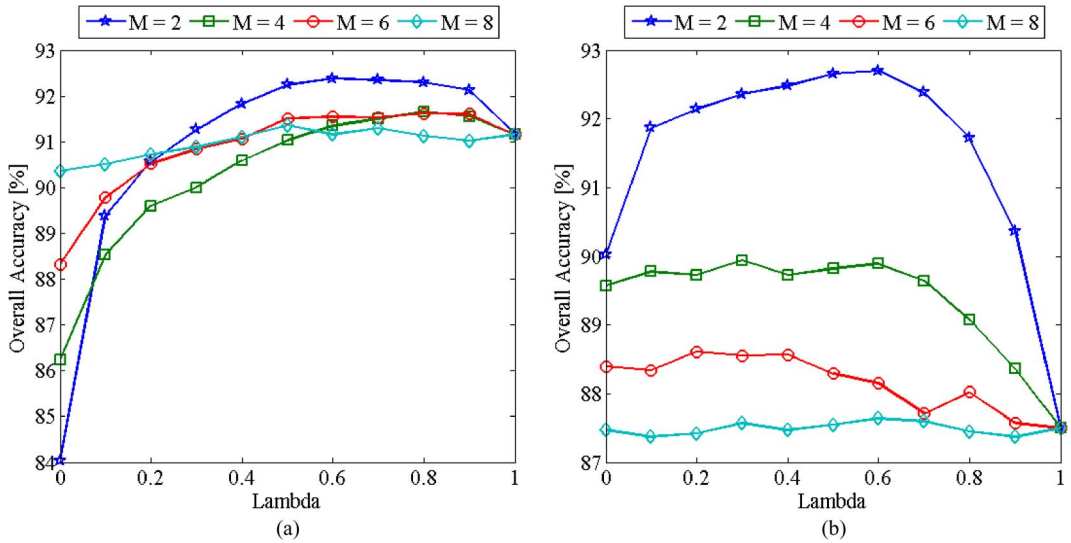

Fig. 7. Classification results obtained by the proposed method after using different values of parameters λ and M for the (a) AVIRIS Indian Pines and (b) ROSIS Pavia University scenes. (a) AVIRIS Indian Pines. (b) ROSIS Pavia University.

described in Section IV-A and bearing in mind the experimental setting described in Section IV-B. The experiments reported in this section can be summarized as follows. In a first experiment, we perform an assessment of the impact of parameter λ, which controls the degree of global information and local information used by the presented method. In a second experiment, we evaluate the impact of parameter M, which controls the number of class combinations for local estimation. In a third experiment, we compare the proposed method with other state-of-the-art methods. Finally, in a fourth experiment, we analyze the sensitivity of the considered method to different training sets and show the good performance of the proposed approach in the presence of limited training samples.

1) Experiment 1. Impact of Parameter λ: In this experiment, we perform an analysis of the impact of parameter λ for the SVM-MLR*sub*-MRF algorithm by using the considered (synthetic and real) data sets with $M = 2$; this means that we set the number of class combinations to 2. The reason for this selection is to keep the number of mixtures per pixel low, as it is often the case in real scenarios. Table I shows the overall (OA) and average (AA) classification accuracies (as a function of parameter λ) for the different scenes considered. For illustrative purposes, Fig. 7 provides a detailed analysis of the classification results obtained by the proposed method after using different values of parameters λ and M for the two considered hyperspectral scenes. In all experiments, we choose 50 random samples per class for training so that, in total, we have 400, 697, and 450 training samples for the synthetic data, AVIRIS Indian pines, and ROSIS Pavia University data sets, respectively (it should be noted that, for the AVIRIS Indian Pines scene, we may select a different number of samples for the small classes).

Several conclusions can be obtained from Table I and Fig. 7. First and foremost, it is remarkable that the proposed approach, which integrates the global information and local information, obtained the best performance in comparison with those results

TABLE II
OVERALL (OA) AND AVERAGE (AA) CLASSIFICATION ACCURACIES (IN PERCENT; AS A FUNCTION OF PARAMETER M, WITH FIXED $\lambda = 0.5$) OBTAINED BY THE SVM-MLR*sub*-MRF METHOD FOR THE SYNTHETIC AND REAL DATA SETS CONSIDERED IN THE EXPERIMENTS. THE BEST RESULTS ARE OUTLINED IN BOLD TYPEFACE

Data	Accuracies	M						
		2	3	4	5	6	7	8
Synthetic	OA	73.61	**74.26**	73.44	72.88	72.16	71.28	70.50
(pure training samples)	AA	72.84	**74.07**	73.43	73.30	72.76	71.89	71.19
Synthetic	OA	79.27	**79.73**	79.32	78.54	77.49	76.08	75.30
(mixed training samples)	AA	79.13	**79.95**	79.46	78.80	77.84	76.55	75.81
AVIRIS Indian Pines	OA	**92.24**	91.60	91.05	90.90	91.47	91.58	91.17
	AA	**95.08**	94.73	94.42	94.06	94.18	94.15	94.06
Pavia University	OA	**92.68**	90.64	89.82	88.93	88.19	87.89	87.58
	AA	**91.93**	90.17	89.45	88.42	88.29	88.11	88.03

TABLE III
OVERALL (OA), AVERAGE (AA), AND INDIVIDUAL CLASS ACCURACIES (IN PERCENT), AND KAPPA STATISTIC (κ) OBTAINED BY DIFFERENT METHODS WITH THE SYNTHETIC IMAGE DATA SET, USING A TOTAL OF 450 TRAINING SAMPLES (50 PER CLASS)

Data	Accuracies	SVM	MLRsub	SVM-MLRsub			SVM-MRF	MLRsub-MRF	SVM-MLRsub-MRF		
				$M=2$	$M=3$	$M=4$			$M=2$	$M=3$	$M=4$
Synthetic	Overall	65.23	68.49	69.77	71.33	71.09	65.03	70.49	73.61	**74.26**	73.44
(pure	Average	66.32	68.97	68.44	70.54	70.65	62.74	71.19	72.84	**74.07**	73.43
training samples)	κ	59.38	62.89	64.09	66.02	65.78	58.23	65.21	68.64	**69.37**	68.57
Synthetic	Overall	69.33	72.18	74.68	75.95	75.70	69.68	75.33	79.27	**79.73**	79.32
(mixed	Average	69.84	72.58	74.08	75.52	75.57	68.68	75.83	79.13	**79.95**	79.46
training samples)	κ	63.88	67.04	69.86	71.35	71.07	63.78	70.72	75.31	**75.90**	75.34

obtained only from the local ($\lambda = 0$) or global ($\lambda = 1$) information. While $\lambda \in [0.4, 0.6]$, the proposed approach obtained very good results for all considered data sets (hence, a reasonable setting is to assign equal weight to local information and global information, i.e., $\lambda = 0.5$). For other suboptimal values of λ, the obtained results are still better or comparable to those obtained by using the local or global information alone.

Furthermore, Fig. 7(a) reveals that the results obtained for the AVIRIS Indian Pines using the global information only ($\lambda = 1$) are better than those obtained using the local information alone ($\lambda = 0$). In turn, Fig. 7(b) reveals an opposite behavior for the ROSIS Pavia University data. However, it is clear from Fig. 7 that an intermediate value of λ (which is equivalent to considering both local and global probabilities) leads to good classification results in the two considered cases, particularly when the value of M is low (this is expected, since the number of mixtures in a given pixel is generally low). It can also be seen in Fig. 7 that the value of M is more relevant for the ROSIS Pavia University data than for the AVIRIS Indian Pines data, which is related with the different spatial resolutions of the considered scenes. From this experiment, we conclude that, in the considered case studies, low values of M and values of λ that ensure a good balance between local information and global information lead to good classification results.

2) Experiment 2. Impact of Parameter M: In this experiment, we perform an evaluation of the impact of parameter M (controlling the number of class combinations) on the presented approach. Table II shows the classification results obtained by the proposed approach using different values of parameter M for all of the considered data sets. For the real data, the best results are obtained with $M = 2$, which means that most of the pixels are either pure or made up of two mixing components. This is a reasonable assumption since, in reality, most pixels are made up by a mixture of a limited number of materials, especially for images with high spatial resolution. However, in our synthetic image experiments, the mixed pixels were simulated in a way that pixels mixed by a higher number of materials are present in the scene. As a result, in the synthetic image experiments, better results can be obtained by using higher values of M.

3) Experiment 3. Comparison With Other Methods: Table III shows a comparison of the classification results obtained by

TABLE IV
OVERALL (OA), AVERAGE (AA), AND INDIVIDUAL CLASS ACCURACIES, AND KAPPA STATISTIC (κ) OBTAINED BY DIFFERENT METHODS WITH THE AVIRIS INDIAN PINES DATA SET, USING A TOTAL OF 697 TRAINING SAMPLES (50 PER CLASS, EXCEPT FOR VERY SMALL CLASSES)

Class	Samples		SVM	MLRsub	SVM-MLRsub			SVM-MRF	MLRsub-MRF	SVM-MLRsub-MRF		
	Train	Test			$M=2$	$M=3$	$M=4$			$M=2$	$M=3$	$M=4$
Alfalfa	27	27	86.42	75.31	79.01	78.64	74.69	93.39	**98.10**	96.30	97.78	97.78
Corn-no till	50	1384	66.45	59.08	68.18	66.01	63.28	79.52	85.80	88.07	**89.46**	88.93
Corn-min till	50	784	68.76	61.27	67.73	61.55	58.78	87.05	89.75	**90.58**	89.13	88.68
Corn	50	184	84.13	79.96	84.20	84.40	83.70	**99.26**	98.88	98.99	98.82	98.89
Grass/trees	50	447	91.35	85.15	88.40	88.86	86.60	**96.73**	93.26	94.50	94.53	93.96
Grass/pasture	50	697	90.82	94.13	96.46	96.39	95.97	98.77	99.18	99.24	**99.27**	99.21
Grass/pasture-mowed	13	13	87.95	65.38	85.13	81.03	68.72	95.90	**98.35**	96.92	96.92	97.44
Hay-windrowed	50	439	95.87	98.73	98.37	98.07	98.56	99.34	99.48	99.45	99.46	**99.51**
Oats	10	10	85.33	83.00	90.33	88.67	87.67	94.52	92.38	**100**	97.00	95.33
Soybeans-notill	50	918	76.32	65.53	70.50	66.80	65.69	91.18	**94.21**	93.97	93.81	93.69
Soybeans-min till	50	2418	62.31	49.41	59.24	52.96	49.14	**90.23**	84.54	86.68	83.67	82.07
Soybeans-clean till	50	564	74.98	77.16	80.67	83.92	84.12	95.22	94.51	97.38	**97.42**	97.33
Wheat	50	162	97.59	99.63	98.93	99.36	99.51	98.90	99.86	**99.90**	**99.90**	99.84
Woods	50	1244	87.08	94.48	94.47	94.67	94.81	88.27	**99.52**	98.04	98.91	99.12
Bldg-Grass-Tree-Drives	50	330	70.87	41.45	53.84	49.10	46.64	**99.70**	77.85	81.78	79.99	79.42
Stone-Steel towers	47	48	98.26	93.89	89.17	92.36	93.54	**99.86**	99.03	99.17	99.58	99.51
OA			75.21	69.32	75.09	72.43	70.56	90.60	91.14	**92.24**	91.60	91.05
AA			82.78	76.47	81.54	80.17	78.21	94.24	94.04	**95.08**	94.73	94.42
κ			71.94	65.34	71.73	68.82	66.78	89.28	89.91	**91.12**	90.42	89.81

the proposed approach (using different values of parameter M) with regard to those provided by other methods for the synthetic image data. Similarly, Tables IV and V show a comparison of the presented approach to other methods using the AVIRIS Indian Pines and ROSIS Pavia University data sets, respectively. In all cases, different values of parameter M were considered for the proposed method, and parameter λ was set to 0.5.

Several conclusions can be obtained from the experiments reported in Tables III–V. First and foremost, it is noticeable that the MLR*sub*-MRF and SVM-MLR*sub*-MRF, which include spatial information and also perform mixed pixel characterization, outperform the SVM-MRF which does not characterize mixed pixels. For instance, Table V reveals that the proposed SVM-MLR*sub*-MRF approach obtained an OA of 92.68%, which contrasts with the OA of 83.96% achieved by SVM-MRF in the ROSIS Pavia University experiments. Similarly, Table III also reveals an OA of about 9.5% larger than that obtained by the SVM-MRF algorithm. However, the situation is different for the experiments with the AVIRIS Indian Pines reported in Table IV. Specifically, the MLR*sub*-MRF did not show a significant improvement with regard to the SVM-MRF, and consequently, the results obtained by the SVM-MLR*sub*-MRF method are not significant. The main reason for this behavior is the difference in the reference data for different classes in the AVIRIS Indian Pines scene, which varies from 20 to 2468 pixels per class. For these data, we chose very limited training sets, and the samples are not evenly distributed among classes. For example, if we take one of the large classes such as *Soybeans-min till* (which contains 2418 samples), we only consider 50 samples for training, which is less than 0.02% of this class. This number of training samples is not enough to completely characterize the class, as it is also the case for other similar classes such as *Soybeans-no till* and *Soybeans-clean till*. Another problem observed in the classification of this image is the fact that class *Bldg-Grass-Tree-Drives* is a highly mixed class. Hence, it is not efficient to use the training samples from this class in the subspace projection procedure, and consequently, the classification accuracies for the methods MLR*sub*-MRF and SVM-MLR*sub*-MRF are not significantly increased with regard to other methods as it was the case in the experiments with other data sets.

If we focus on the results reported in Table III in a fully controlled environment, we can conclude that the class probabilities estimated by the SVM classifier may not be fully

TABLE V
OVERALL (OA), AVERAGE (AA), AND INDIVIDUAL CLASS ACCURACIES (IN PERCENT), AND KAPPA STATISTIC (κ) OBTAINED BY DIFFERENT METHODS WITH THE ROSIS PAVIA UNIVERSITY DATA SET, USING A TOTAL OF 450 TRAINING SAMPLES (50 PER CLASS)

Class	Samples		SVM	MLRsub	SVM-MLRsub			SVM-MRF	MLRsub-MRF	SVM-MLRsub-MRF		
	Train	Test			$M=2$	$M=3$	$M=4$			$M=2$	$M=3$	$M=4$
Asphalt	50	6631	73.67	43.99	65.73	59.55	54.52	**95.05**	80.13	91.05	89.79	85.62
Meadows	50	18649	63.79	66.31	75.95	72.55	70.90	69.63	89.71	**94.26**	93.00	93.16
Gravel	50	2099	71.34	63.80	67.07	63.70	63.55	72.61	71.56	**73.66**	68.33	68.42
Trees	50	3064	96.55	79.96	89.70	86.98	84.44	**97.93**	80.45	93.83	92.24	88.82
Meta sheets	50	1345	99.35	99.08	98.88	99.19	99.24	99.87	99.70	99.58	99.77	**99.78**
Bare soil	50	5029	87.06	63.96	65.97	61.81	60.83	**98.19**	91.09	91.55	84.34	84.25
Bitumen	50	1330	91.62	86.08	83.91	83.74	84.24	**95.86**	90.43	86.63	89.06	90.44
Bricks	50	3682	85.21	62.64	74.17	69.57	67.19	**97.92**	91.03	96.90	95.12	94.70
Shadows	50	947	98.90	98.52	99.10	98.61	98.50	99.13	99.72	**99.86**	99.85	99.83
OA			75.38	65.47	75.07	71.38	69.39	83.96	87.50	**92.68**	90.64	89.82
AA			85.28	73.81	80.05	77.30	75.94	91.80	88.20	**91.93**	90.17	89.45
κ			69.49	56.74	68.06	63.59	61.18	79.96	83.72	**90.40**	87.71	86.62

TABLE VI
STATISTICAL SIGNIFICANCE OF THE DIFFERENCES IN CLASSIFICATION ACCURACIES (MEASURED USING THE MCNEMAR'S TEST IN [50]) FOR DIFFERENT METHODS WITH ALL OF THE CONSIDERED DATA SETS

Hyperspectral data set	Value of Z calculated by the McNemar's test			
	SVM-MLRsub/SVM	SVM-MLRsub/MLRsub	SVM-MLRsub-MRF/SVM-MRF	SVM-MLRsub-MRF/MLRsub-MRF
Synthetic (Pure samples)	9.87	4.44	17.33	8.58
Synthetic (Mixed samples)	11.58	7.33	19.33	11.19
AVIRIS Indian Pines	-3.77	18.98	1.07	16.46
ROSIS Pavia University	-1.19	48.72	44.10	33.69

reliable to be used in the MRF regularization procedure. This is because of the nature of the SVM method, which is a hard classifier. In turn, the MLR*sub*-MRF method better characterized noise and mixed pixels. However, the SVM-MLR*sub*-MRF method provided the highest classification accuracies in this experiment. For instance, in the synthetic experiment using pure training samples, the OA achieved by the presented method improved by 3.77% and 9.23% of the OA achieved by the MLR*sub*-MRF and SVM-MRF, respectively. When mixed training samples were used, the proposed SVM-MLR*sub*-MRF algorithm obtained an OA of 79.73%, which is the best result for this data set (4.40% and 10.05% higher than MLR*sub*-MRF and SVM-MRF, respectively). This is because mixed training samples are near the decision boundaries and can be very effective in class discrimination.

If we now focus on Table V, we can observe that, in this experiment, the pixelwise SVM classifier already provides high classification accuracies. However, including the spatial–contextual information significantly improves the classification accuracies as it can be particularly observed in the SVM-MLR*sub*-MRF method. Here, by using only 50 training samples per class (in total 450 samples, which is a very low number for this scene), the proposed algorithm obtained an OA of 92.68%, which is 5.18% and 8.72% higher than MLR*sub*-MRF and SVM-MRF, respectively.

In order to analyze the statistical significance of the results obtained by the different compared methods, we have used McNemar's test [50]. In this test, a value of $|Z| > 1.96$ indicates that there is a significant difference in accuracy between two classification methods. The sign of Z is also a criterion to indicate whether a first classifier is more accurate than a second one ($Z > 0$) or vice versa ($Z < 0$). Table VI provides the results obtained for all of the considered (synthetic and real) data sets. As it can be seen from Table VI, the differences in classification accuracies between our proposed method (implemented with parameters $M = 2$ and $\lambda = 0.5$) and the MLR*sub*-MRF method are statistically significant. Compared to the SVM-MRF method, SVM-MLR*sub*-MRF exhibits a statistically

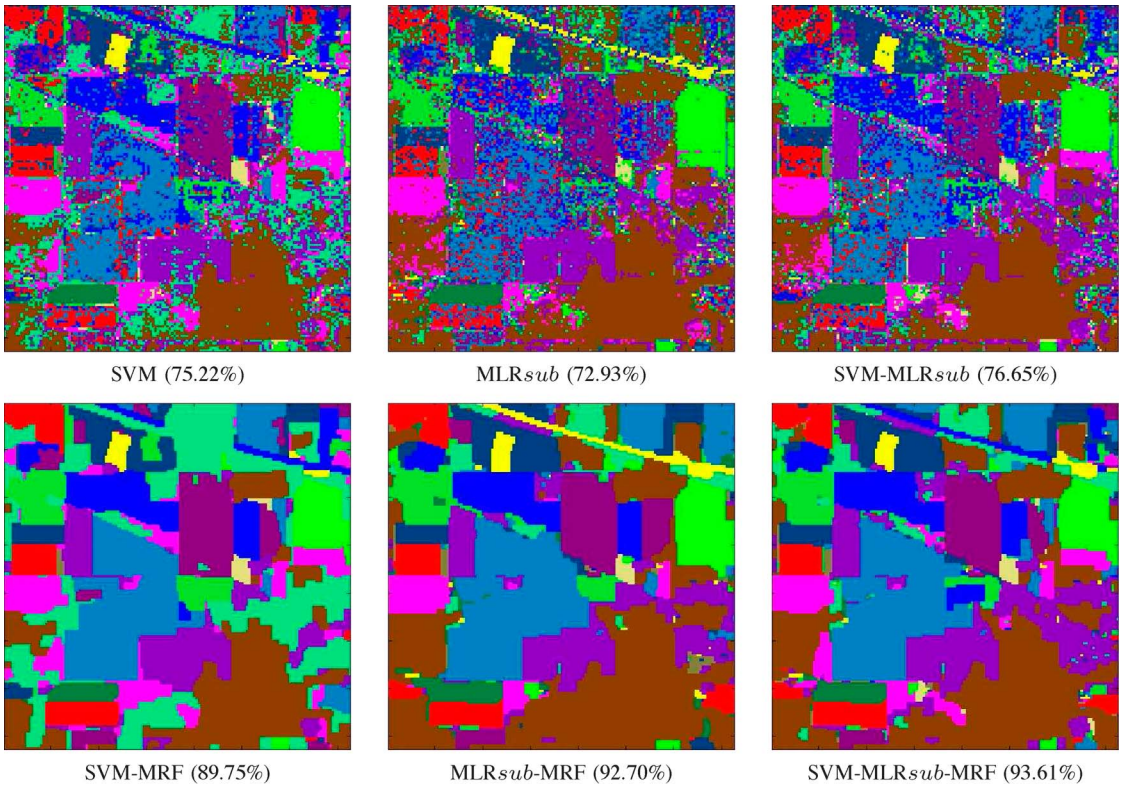

Fig. 8. Classification results and overall classification accuracies (in parentheses) obtained by different methods for the AVIRIS Indian Pines data set.

significant improvement in classification accuracies except for the AVIRIS Indian Pines scene, in which the McNemar's test indicates that the performances of the two methods are similar.

For illustrative purposes, Figs. 8 and 9 show some of the obtained classification maps for the AVIRIS Indian Pines and ROSIS Pavia University data sets, respectively. Each of the maps corresponds to one out of the 30 Monte Carlo experiments which were averaged to produce the results, respectively, reported in Tables IV and V. As shown in Figs. 8 and 9, good classification results with adequate class delineation and spatial consistency can be observed for the presented method (which was run in both cases using $M = 2$ and $\lambda = 0.5$) in comparison to other approaches.

4) Experiment 4. Impact of the Number of Training Samples: In this experiment, we first conduct an evaluation of the impact of the number of training samples on the proposed approach, using the two real data sets and fixing parameters $M = 2$ and $\lambda = 0.5$. Table VII shows the classification accuracies obtained by the presented method as a function of the number of training samples per class (where the total number of training samples is given in parentheses). In the case of the ROSIS Pavia University scene, we also performed an experiment using all of the available training samples (3921) in order to establish an upper bound to the presented results and to explore if, with a more limited training set, the results are indeed close to that bound.

The results reported in Table VII show that, for the AVIRIS Indian Pines data, we only need 40 training samples per class (for a total of 570) in order to achieve an OA of around 90.00% (and AA larger than 93%). For the ROSIS Pavia University data, we only need less than 40 samples per class in order to obtain an OA of 90%. This is remarkable, as sometimes it is very difficult to collect large training sets in practice. Table VII also reveals that the presented method provided results which are comparable to those provided by the SVM-MRF and superior than those provided by MLR*sub*-MRF with the AVIRIS Indian Pines data. For the ROSIS Pavia University data, the proposed approach obtained an OA of 94.57% using only 80 training samples per class (for a total of 720 samples). This result is quite close to the upper bound result, obtained using all available (3921) training samples for this scene. In fact, this result is 4.32% higher than the results provided by MLR*sub*-MRF and 9.89% higher than the results provided by the SVM-MRF. This leads to two main observations. First, by including the local information, the SVM-MLR*sub*-MRF greatly improved the performance obtained by the MLR*sub*-MRF algorithm which only considers the global information. A second observation is that the methods that characterize mixed pixels, i.e., MLR*sub*-MRF and SVM-MLR*sub*-MRF, can outperform the methods that do not incorporate mixed pixel characterization, even for scenes collected at high spatial resolution.

To conclude this section, Fig. 10 compares the performances of the methods: SVM-MRF, MLR*sub*-MRF, and SVM-MLR*sub*-MRF in 30 Monte Carlo runs conducted for different random sets of training sample sets for the two real hyperspectral scenes (AVIRIS Indian Pines, at the top of the figure, and ROSIS Pavia University, at the bottom of the figure). In order to establish a fair comparison, in each iteration, the same training set is used by all three methods. As Fig. 10 shows, when compared with MLR*sub*-MRF, the proposed method shows more uniform results and appears less sensitive to the quality of

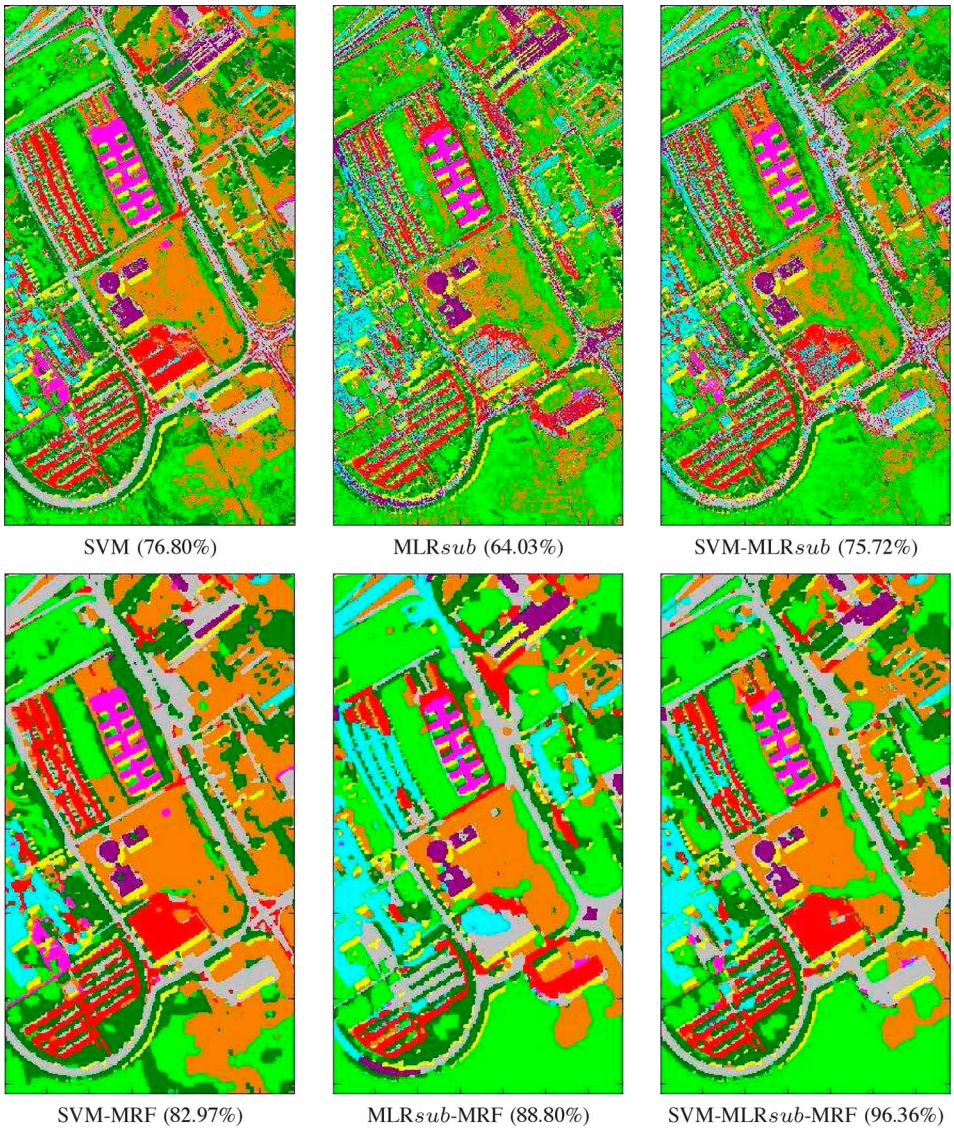

Fig. 9. Classification results and overall classification accuracies (in parentheses) obtained by different methods for the ROSIS Pavia University data set.

training samples. When compared with SVM-MRF, the proposed method shows slightly superior results for the AVIRIS Indian Pines scene and consistently better performance for the Pavia University scene. Again, we reiterate that the SVM-MLR*sub* method takes the advantages of both SVM and MLR*sub* and can compensate the situation in which one of the methods does not provide good performance by taking advantage of the other method. This is also the reason why the proposed SVM-MLR*sub*-MRF method can provide good performance in those cases in which none of the methods SVM-MRF and MLR*sub*-MRF exhibits good classification accuracies. This is the case, for instance, in iterations 6, 13, and 25 for the ROSIS Pavia University experiments reported in Fig. 10.

V. Conclusion and Future Research Lines

In this paper, we have introduced a novel spectral–spatial classifier for hyperspectral image data. The proposed method is based on the consideration of both global posterior probability distributions and local probabilities which result from the whole image and a set of previously derived class combination maps, respectively. The proposed approach, which intends to characterize mixed pixels in the scene and assumes that these pixels are normally mixed by only a few components, provides some distinguishing features with regard to other existing approaches. At the local learning level, the presented method removes the impact of irrelevant classes by means of a preprocessing stage (implemented using the probabilistic SVM) intended to produce a subset of M most probable classes for each pixel.

This stage locally eliminates noise and enhances the impact of the most relevant classes. These aspects, together with the joint characterization of mixed pixels and spatial–contextual information, make our method unique and representative of a framework that, for the first time in the literature, integrates local and global probabilities in the analysis of hyperspectral data in order to constrain the number of mixing components used in the characterization of mixed pixels. This is consistent with the observation that, despite the presence of mixed pixels in real hyperspectral scenes, it is reasonable to assume that the

TABLE VII
OVERALL (OA) AND AVERAGE (AA) ACCURACY (IN PERCENT) AS A FUNCTION OF THE NUMBER OF TRAINING SAMPLES PER CLASS FOR THE SVM-MLR*sub*-MRF METHOD, WHERE THE TOTAL NUMBER OF TRAINING SAMPLES IS GIVEN IN PARENTHESES

Methods	Accuracies	AVIRIS Indian Pines				ROSIS Pavia University				
		20 (303)	40 (570)	60 (817)	80 (1057)	20 (180)	40 (360)	60 (540)	80 (720)	All samples (3921)
SVM	OA	65.67	72.48	76.69	78.85	70.02	73.70	75.08	77.21	81.13
	AA	75.98	81.01	83.26	84.66	80.50	84.24	85.65	86.71	89.05
MLRsub	OA	65.36	68.16	69.72	70.07	64.84	64.89	64.70	65.10	70.61
	AA	75.10	76.10	75.65	75.10	72.72	74.08	73.97	74.16	73.92
SVM-MLRsub	OA	68.29	73.39	76.38	78.00	71.35	73.96	75.29	76.69	82.61
	AA	76.44	80.42	82.15	82.43	77.50	80.03	80.92	81.61	83.80
SVM-MRF	OA	82.10	88.65	91.17	92.01	79.56	82.93	82.54	84.68	86.17
	AA	89.54	93.52	94.26	94.36	86.13	91.12	91.68	92.36	92.56
MLRsub-MRF	OA	77.47	87.80	91.41	92.83	83.50	87.18	89.08	90.25	93.10
	AA	85.84	92.02	94.23	95.31	84.73	88.61	89.75	89.44	86.67
SVM-MLRsub-MRF	OA	82.70	89.79	93.10	93.86	85.88	90.97	92.10	94.57	95.56
	AA	88.89	93.40	95.52	96.14	87.06	91.17	92.70	93.06	90.69

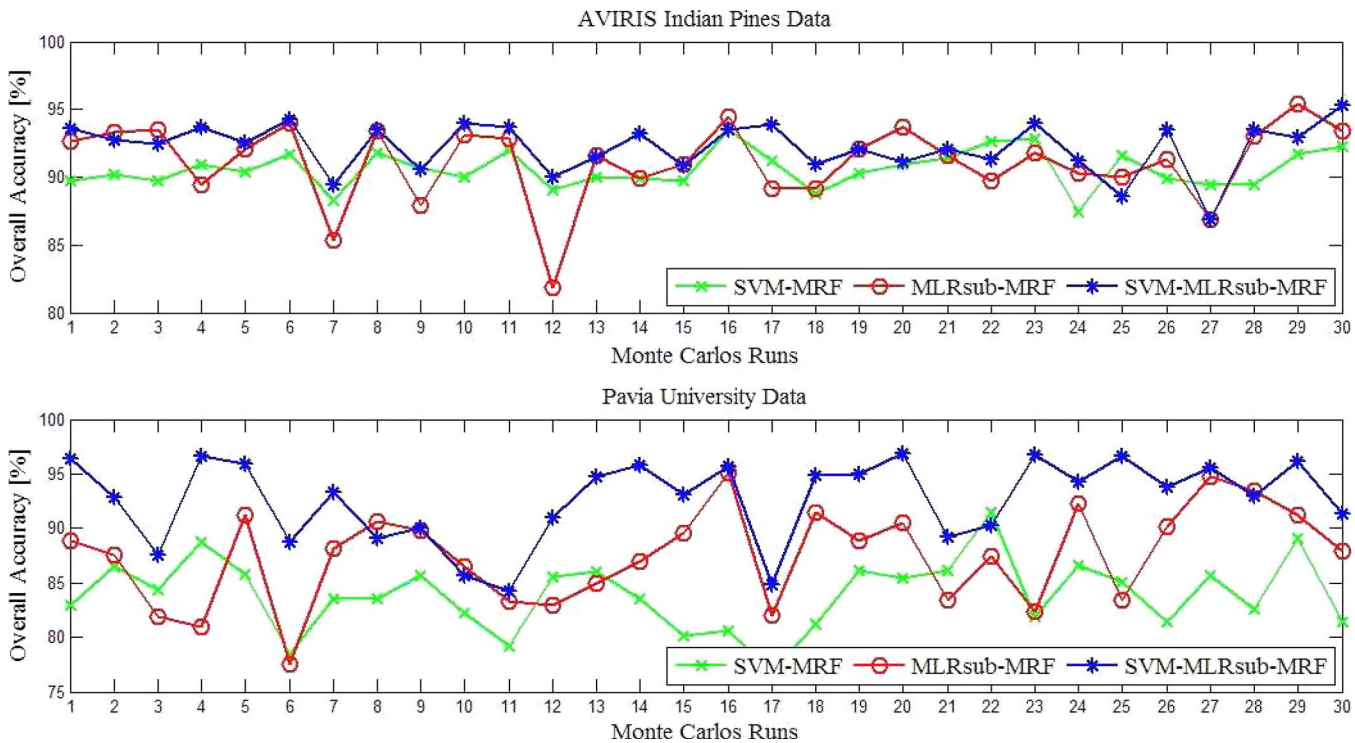

Fig. 10. Comparison of the performances of the methods: SVM-MRF, MLR*sub*-MRF, and SVM-MLR*sub*-MRF in 30 Monte Carlo runs conducted for different random sets of training sample sets for the two real hyperspectral scenes: (top) AVIRIS Indian Pines and (bottom) ROSIS Pavia University. In each run, the same training set is used by all three methods.

mixing components in a given pixel are limited. Our experimental results, conducted using both synthetic and real hyperspectral scenes widely used in the hyperspectral classification community, indicate that the proposed approach leads to state-of-the-art performance when compared with other approaches, particularly in scenarios in which very limited training samples are available.

As future research, we are currently developing a version of the presented algorithm in which parameter M is adaptively estimated for each pixel rather than set in advance as in the

version of the algorithm reported in this paper. Interestingly, we have empirically observed that the adaptive selection produces similar results to those obtained in this work with fixed parameter settings such as $M = 2$ or $M = 3$, which result in much lower computational cost than an adaptive estimation of the parameter on a per-pixel basis. As a result, we will continue exploring the possibility to select this parameter adaptively in order to improve the obtained classification results without increasing computational complexity, which currently stays on the same order of magnitude as the other methods used in the comparisons reported in this work. In future developments, we will further explore the relationship between the parameters of our method and the spatial resolution, level of noise, and complexity of the analyzed scenes. We are also planning on exploring the applications of the presented method for the analysis of multitemporal data sets.

ACKNOWLEDGMENT

The authors would like to thank Prof. D. Landgrebe for making the AVIRIS Indian Pines hyperspectral data set available to the community, Prof. P. Gamba for providing the ROSIS data over University of Pavia, Pavia, Italy, along with the training and test sets, and the Associate Editor who handled this paper and the three anonymous reviewers for providing truly outstanding comments and suggestions that significantly helped in improving the technical quality and presentation of this paper.

REFERENCES

[1] C.-I. Chang, *Hyperspectral Data Exploitation: Theory and Applications*. New York, NY, USA: Wiley, 2007.

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

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

[4] A. Plaza, J. A. Benediktsson, J. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, J. A. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, no. S1, pp. 110–122, Sep. 2009.

[5] R. A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing,* 3rd ed. New York, NY, USA: Academic, 2007.

[6] L. Samaniego, A. Bardossy, and K. Schulz, "Supervised classification of remotely sensed imagery using a modified k-NN technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2112–2125, Jul. 2008.

[7] S. Subramanian, N. Gat, M. Sheffield, J. Barhen, and N. Toomarian, "Methodology for hyperspectral image classification using novel neural network," in *Proc. SPIE Algorithms Multispectr. Hyperspectr. Imagery III*, Aug. 1997, vol. 3071, pp. 128–137.

[8] H. Yang, F. V. D. Meer, W. Bakker, and Z. J. Tan, "A back-propagation neural network for mineralogical mapping from AVIRIS data," *Int. J. Remote Sens.*, vol. 20, no. 1, pp. 97–110, Jan. 1999.

[9] C. Hernández-Espinosa, M. Fernández-Redondo, and J. Torres-Sospedra, "Some experiments with ensembles of neural networks for classification of hyperspectral images," in *Proc. ISNN*, 2004, vol. 1, pp. 912–917.

[10] G. M. Foody and A. Mathur, "The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM," *Remote Sens. Environ.*, vol. 103, no. 2, pp. 179–189, Jul. 2006.

[11] J. A. Gualtieri and R. F. Cromp, "Support vector machines for hyperspectral remote sensing classification," in *Proc. SPIE*, 1998, vol. 4584, pp. 506–508.

[12] C. Huang, L. S. Davis, and J. R. Townshend, "An assessment of support vector machines for land cover classification," *Int. J. Remote Sens.*, vol. 23, no. 4, pp. 725–749, Feb. 2002.

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

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

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

[16] A. Villa, J. Chanussot, J. A. Benediktsson, and C. Jutten, "Spectral unmixing for the classification of hyperspectral images at a finer spatial resolution," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 521–533, Jun. 2011.

[17] X. Jia, C. Dey, D. Fraser, L. Lymburner, and A. Lewis, "Controlled spectral unmixing using extended support vector machines," in *Proc. 2nd WHISPERS*, Reykjavik, Iceland, Jun. 2010, pp. 1–4.

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

[19] P. C. Smits, "Multiple classifier systems for supervised remote sensing image classification based on dynamic classifier selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 4, pp. 801–813, Apr. 2002.

[20] J. A. Benediktsson, J. Chanussot, and M. Fauvel, "Multiple classifier systems in remote sensing: From basics to recent developments," in *Proc. Multiple Classif. Syst.*, 2007, vol. 4472, pp. 501–512.

[21] G. M. Foody, D. S. Boyd, and C. Sanchez-Hernandez, "Mapping a specific class with an ensemble of classifiers," *Int. J. Remote Sens.*, vol. 28, no. 8, pp. 1733–1746, Apr. 2007.

[22] P. Du, J. Xia, W. Zhang, K. Tan, Y. Liu, and S. Liu, "Multiple classifier system for remote sensing image classification: A review," *Sensors*, vol. 12, no. 4, pp. 4764–4792, Apr. 2012.

[23] J. Ham, Y. Chen, and M. M. Crawford, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 492–501, Mar. 2005.

[24] J. C.-W. Chan and D. Paelinckx, "Evaluation of random forest and adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery," *Remote Sens. Environ.*, vol. 112, no. 6, pp. 2299–3011, Jun. 2008.

[25] S. Kumar, J. Ghosh, and M. M. Crawford, "Hierarchical fusion of multiple classifiers for hyperspectral data analysis," *Pattern Anal. Appl.*, vol. 5, no. 2, pp. 210–220, Jun. 2002.

[26] X. Ceamanos, B. Waske, J. A. Benediktsson, J. Chanussot, M. Fauvel, and J. R. Sveinsson, "A classifier ensemble based on fusion of support vector machines for classifying hyperspectral data," *Int. J. Image Data Fusion*, vol. 1, no. 4, pp. 293–307, Dec. 2010.

[27] R. Kettig and D. Landgrebe, "Classification of multispectral image data by extraction and classification of homogenous objects," *IEEE Trans. Geosci. Electron.*, vol. GE-14, no. 1, pp. 19–26, Jan. 1976.

[28] H. Ghassemian and D. Landgrebe, "Object-oriented feature extraction method for image data compaction," *IEEE Control Syst. Mag.*, vol. 8, no. 3, pp. 42–48, Jun. 1988.

[29] S. M. de Jong and F. D. van der Meer, *Remote Sensing Image Analysis: Including the Spatial Domain*. Norwell, MA, USA: Kluwer, 2004.

[30] Y. Wang, R. Niu, and X. Yu, "Anisotropic diffusion for hyperspectral imagery enhancement," *IEEE Sensors J.*, vol. 10, no. 3, pp. 469–477, Mar. 2010.

[31] S. Velasco-Forero and V. Manian, "Improving hyperspectral image classification using spatial preprocessing," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 2, pp. 297–301, Apr. 2009.

[32] van der Linden, A. Janz, B. Waske, M. Eiden, and P. Hostert, "Classifying segmented hyperspectral data from a heterogeneous urban environment using support vector machines," *J. Appl. Remote Sens.*, vol. 1, no. 1, p. 013543, Mar.–Oct. 2007.

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

[34] S. Z. Li, *Markov Random Field Modeling in Image Analysis,* 3rd ed. London, U.K.: Springer-Verlag, 2009.

[35] A. Farag, R. Mohamed, and A. El-Baz, "A unified framework for map estimation in remote sensing image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 7, pp. 1617–1634, Jul. 2005.

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

[37] 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.
[38] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.
[39] T.-F. Wu, C.-J. Lin, and R. C. Weng, "Probability estimates for multiclass classification by pairwise coupling," *J. Mach. Learn. Res.*, vol. 5, pp. 975–1005, Dec. 2004.
[40] V. Vapnik and A. Chervonenkis, "The necessary and sufficient conditions for consistency in the empirical risk minimization method," *Pattern Recognit. Image Anal.*, vol. 1, no. 3, pp. 283–305, 1991.
[41] C.-J. Lin, H.-T. Lin, and R. C. Weng, "A note on Platt's probabilistic outputs for support vector machines," Dept. Comput. Sci., Nat. Taiwan Univ., Taipei, Taiwan, 2003.
[42] C. Chang and C. Lin, LIBSVM: A Library for Support Vector Machines 2009. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm/
[43] D. Böhning, "Multinomial logistic regression algorithm," *Ann. Inst. Stat. Math.*, vol. 44, no. 1, pp. 197–200, Mar. 1992.
[44] P. Clifford, "Markov random fields in statistics," in *Disorder in Physical Systems: A Volume in Honour of John M. Hammersley*. Oxford, U.K.: Clarendon, 1990, pp. 19–32.
[45] Y. Boykov, O. Veksler, and R. Zabih, "Efficient approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.
[46] J. A. Benediktsson and P. H. Swain, "Consensus theoretic classification methods," *IEEE Trans. Syst., Man, Cybern.*, vol. 22, no. 4, pp. 688–704, Jul./Aug. 1992.
[47] J. A. Benediktsson and I. Kanellopoulos, "Classification of multisource and hyperspectral data based on decision fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1367–1377, May 1999.
[48] R. O. Green, M. L. Eastwood, C. M. Sarture, T. G. Chrien, M. Aronsson, B. J. Chippendale, J. A. Faust, B. E. Pavri, C. J. Chovit, M. Solis, M. R. Olah, and O. Williams, "Imaging spectroscopy and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)," *Remote Sens. Environ.*, vol. 65, no. 3, pp. 227–248, Sep. 1998.
[49] J. Nascimento and J. Bioucas-Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.
[50] G. M. Foody, "Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 5, pp. 627–633, May 2004.

Mahdi Khodadadzadeh (S'10) received the B.Sc. degree in electrical engineering from the Sadjad Institute of Higher Education, Mashhad, Iran, in 2008 and the M.Sc. degree from Tarbiat Modares University, Tehran, Iran, in 2011. He is currently working toward the Ph.D. degree in the Hyperspectral Computing Laboratory (HyperComp), Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, Cáceres, Spain.

His research interests include remote sensing, pattern recognition, and signal and image processing, with particular emphasis on spectral and spatial techniques for hyperspectral image classification.

Mr. Khodadadzadeh is a manuscript reviewer of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.

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

From 2007 to 2011, she was a Marie Curie Research Fellow with the Departamento de Engenharia Electrotécnica e de Computadores and the Instituto de Telecomunicações, IST, Universidade Técnica de Lisboa, in the framework of the European Doctorate for Signal Processing (SIGNAL). She has also been actively involved in the Hyperspectral Imaging Network, a Marie Curie Research Training Network involving 15 partners in 12 countries and intended to foster research, training, and cooperation on hyperspectral imaging at the European level. Since 2011, she has been a Postdoctoral Researcher with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, Cres, Spain. She has been a Reviewer of several journals, including the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, *Pattern Recognition*, *Optical Engineering*, *Journal of Applied Remote Sensing*, and *Inverse Problems and Imaging*. Her research interests include hyperspectral image classification and segmentation, spectral unmixing, signal processing, and remote sensing.

Dr. Li received the 2012 Best Reviewer Award of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.

Antonio Plaza (M'05–SM'07) is an Associate Professor (with accreditation for Full Professor) with the Department of Technology of Computers and Communications, University of Extremadura, Cres, Spain, where he is the Head of the Hyperspectral Computing Laboratory (HyperComp). He was the Coordinator of the Hyperspectral Imaging Network, a European project with a total funding of 2.8 MEuro (2007–2011). He is the author of more than 370 publications, including more than 100 JCR journal papers (60 in IEEE journals), 20 book chapters, and over 230 peer-reviewed conference proceeding papers (90 in IEEE conferences). He has been the Guest Editor of seven special issues of JCR journals (three in IEEE journals).

Prof. Plaza was the Chair of the IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing in 2011. He was 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, a journal for which he served as Associate Editor in 2007–2012. He is also an Associate Editor of IEEE ACCESS and IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE. He was a member of the Editorial Board of the IEEE GEOSCIENCE AND REMOTE SENSING NEWSLETTER in 2011–2012 and a member of the steering committee of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING in 2012. He served as the Director of Education Activities of the IEEE Geoscience and Remote Sensing Society (GRSS) in 2011–2012, and he has been the President of the Spanish Chapter of IEEE GRSS since November 2012. He has been the Editor-in-Chief of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING since January 2013.

Hassan Ghassemian (M'99–SM'06) was born in Iran in 1956. He received the B.S.E.E. degree from Tehran College of Telecommunication, Tehran, Iran, in 1980 and the M.S.E.E. and Ph.D. degrees from Purdue University, West Lafayette, IN, USA, in 1984 and 1988, respectively.

Since 1988, he has been with Tarbiat Modares University, Tehran, Iran, where he is a Professor of electrical and computer engineering. He has published more than 300 articles in peer-reviewed journals and conference papers. He has trained more than 100 M.S. and Ph.D. students who have assumed key positions in software and computer system design applications related to signal and image processing in the past 25 years. His research interests focus on multisource signal/image processing, information analysis, and remote sensing.

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

Since 1995, he has been with the Department of Electrical and Computer Engineering, IST, where he was an Assistant Professor from 1995 to 2007 and where he has been an Associate Professor since 2007. Since 1993, he has also been a Senior Researcher with the Pattern and Image Analysis Group, Instituto de Telecomunicações, which is a private nonprofit research institution. His research interests include inverse problems, signal and image processing, pattern recognition, optimization, and remote sensing.

Dr. Bioucas-Dias was an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS (1997–2000). He is an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING and IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He was a Guest Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING for the Special Issue on Spectral Unmixing of Remotely Sensed Data and of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING for the Special Issue on Hyperspectral Image and Signal Processing. He is a Guest Editor of the IEEE SIGNAL PROCESSING MAGAZINE for the Special Issue on Signal and Image Processing in Hyperspectral Remote Sensing. He was the General Cochair of the 3rd IEEE GRSS Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing (WHISPERS'2011), and he has been a member of program/technical committees of several international conferences.

Xia Li received the B.S. and M.S. degrees from Peking University, Beijing, China, and the Ph.D. degree in geographical information sciences from the University of Hong Kong, Hong Kong.

He is a Professor and the Director of the Centre for Remote Sensing and Geographical Information Sciences, School of Geography and Planning, Sun Yat-Sen University, Guangzhou, China. He is also a Guest Professor with the Department of Geography, University of Cincinnati, Cincinnati, OH, USA. He got the ChangJiang scholarship and the award of the Distinguished Youth Fund of NSFC in 2005. He is currently on the editorial boards of several international journals including *International Journal of Geographical Information Science*, *Computers, Environment and Urban Systems*, and *GeoJournal*. He is an Associate Editor of a Chinese journal *Tropical Geography*. He has about 200 articles, of which many appeared in international journals. His papers are widely published in top international GIS and remote sensing journals, such as *Remote Sensing of Environment*, *International Journal of Remote Sensing*, *Photogrammetric Engineering & Remote Sensing*, *International Journal of Geographical Information Science*, *Environment and Planning A*, and *Environment and Planning B*. His major research interests include the development of urban cellular automata and agent-based models for simulating urban growth and land use changes. Some of his researches focus on the development of methodologies for calibrating these simulation models. Recently, he has carried out the researches on using ant intelligence for spatial optimization. His other researches include the use of radar remote sensing for urban applications.