Spectral–Spatial Classification of Hyperspectral Data Using Loopy Belief Propagation and Active Learning

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

***Abstract*—In this paper, we propose a new framework for spectral–spatial classification of hyperspectral image data. The proposed approach serves as an engine in the context of which active learning algorithms can exploit both spatial and spectral information simultaneously. An important contribution of our paper is the fact that we exploit the marginal probability distribution which uses the whole information in the hyperspectral data. We learn such distributions from both the spectral and spatial information contained in the original hyperspectral data using loopy belief propagation. The adopted probabilistic model is a discriminative random field in which the association potential is a multinomial logistic regression classifier and the interaction potential is a Markov random field multilevel logistic prior. Our experimental results with hyperspectral data sets collected using the National Aeronautics and Space Administration's Airborne Visible Infrared Imaging Spectrometer and the Reflective Optics System Imaging Spectrometer system indicate that the proposed framework provides state-of-the-art performance when compared to other similar developments.**

***Index Terms*—Active learning (AL), discriminative random fields (DRFs), hyperspectral image classification, loopy belief propagation (LBP), Markov random fields (MRFs), spectral–spatial analysis.**

I. Introduction

REMOTELY sensed hyperspectral imaging instruments are capable of collecting hundreds of images, corresponding to different wavelength channels, for the same area on the surface of the Earth [1]. The concept of hyperspectral imaging was first introduced at National Aeronautics and Space Administration (NASA)'s Jet Propulsion Laboratory [2], where a system called Airborne Imaging Spectrometer was built to demonstrate the concept. Today, NASA is continuously gathering high-dimensional image data with instruments such as Jet Propulsion Laboratory's Airborne Visible–Infrared Imaging Spectrometer (AVIRIS) [3]. This advanced sensor for Earth observation records the visible and near-infrared spectra of the reflected light using more than 200 spectral bands, thus producing a stack of images in which each pixel (vector) is represented by a spectral signal that uniquely characterizes the underlying objects.

The number and variety of processing tasks in hyperspectral remote sensing are enormous [4]. One of the most popular ones is classification, which consists in assigning a label to each pixel in order to generate a thematic land-cover map [5]. The problem of hyperspectral image classification has been tackled in the past using several different approaches. For instance, several machine learning and image processing techniques have been applied to extract relevant information from hyperspectral data during the last decade [6]. The high dimensionality of hyperspectral data in the spectral domain poses critical problems for supervised algorithms [6], [7]; most notably, in order for supervised classifiers to perform properly, there is a need for large training sets in order to avoid the well-known Hughes effect [8], [9]. However, training samples are limited, expensive, and quite difficult to obtain in real remote sensing scenarios. Inspired by this motivation, active learning (AL), which aims at finding the most informative training set, has become an active research topic in recent years [10]–[17].

AL has been adopted in the remote sensing community as an effective strategy to reduce the cost of acquiring large labeled training sets. It is a method of online learning, where a learner strategically selects new training examples that provide maximal information about the unlabeled data set, resulting in higher classification accuracy for a given training set size as compared to using randomly selected examples. In [18], a survey of AL approaches was presented in the context of supervised remote sensing classification problems. In the literature, most available AL algorithms acquire new labels based on spectral information alone, discarding the spatial information present in the original data. For instance, in [10], an AL technique that efficiently updates existing classifiers by using fewer labeled data points than semisupervised methods is presented. That technique is well suited for learning or adapting classifiers when there is a substantial change in the spectral signatures between labeled and unlabeled data. In [11], two AL algorithms for semiautomatic definition of training samples in remote sensing image classification are presented and applied to a variety of remote sensing data, including very high resolution and hyperspectral images. In [12], an efficient AL algorithm with knowledge transfer for hyperspectral data analysis is presented. In [13], an unbiased AL heuristic that enforces diverse sampling is

Manuscript received November 25, 2011; revised May 6, 2012; accepted June 10, 2012. This work was supported in part by the European Community's Marie Curie Research Training Networks Programme under Contract MRTN-CT-2006-035927 (Hyperspectral Imaging Network), by the Portuguese Science and Technology Foundation, under project PEst-OE/EEI/LA0008/2011, and by the Spanish Ministry of Science and Innovation (Calibration of Earth Observation Satellites in SPAIN (CEOS-SPAIN) project, reference AYA2011-29334-C02-02).

J. Li and A. Plaza are with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica de Cáceres, University of Extremadura, 10071 Cáceres, Spain (e-mail: junli@unex.es; aplaza@unex.es).

J. M. Bioucas-Dias is with the Departamento de Engenharia Electrotécnica e de Computadores and the Instituto de Telecomunicações, Instituto Superior Técnico, Universidade Técnica de Lisboa, 1049-1 Lisboa, Portugal (e-mail: bioucas@lx.it.pt).

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.2012.2205263

investigated. In [14], different batch-mode AL techniques for the classification of remote sensing images with support vector machines (SVMs) are discussed. In [15], a new AL technique is proposed to identify the most uncertain samples whose labeling and inclusion in the training set involve a high probability to improve the classification results. In [16], a novel coregularization framework for AL is proposed that explores the intrinsic multiview information embedded in the hyperspectral data by focusing only on samples with high uncertainty. This approach builds a so-called contention pool which is a small subset of the overall unlabeled data pool, thus reducing computational complexity. In [17], a novel batch-mode AL technique is proposed in which the uncertainty of each unlabeled sample is measured by defining a criterion which not only considers the smallest distance to the decision hyperplanes but also takes into account the distances to other hyperplanes if the sample is within the margin of their decision boundaries. Despite the fact that several of the aforementioned AL algorithms mainly focus on analyzing spectral properties, spatial information plays a very important role in the classification of hyperspectral data as it has been shown in previous work [6]. Thus, the combination of spectral and spatial information for AL represents a novel and promising contribution yet to be explored in the hyperspectral imaging literature.

In this paper, we propose a new framework which exploits spatial information in the context of hyperspectral image classification. The proposed framework, based on the marginal probability distribution which is obtained from the whole information in the hyperspectral data, serves as an engine in which AL techniques can exploit both the spectral and the spatial information in the data. Here, the posterior class probability is modeled with a discriminative random field (DRF) [19] in which the association potential is linked with a multinomial logistic regression (MLR) classifier [20], [21], where the regressors are inferred by the logistic regression via variable splitting and augmented Lagrangian algorithm (LORSAL) [22], and the interaction potential modeling the spatial information is linked to a Markov random field (MRF) [23] multilevel logistic (MLL) prior [24]. The training set is enlarged with new samples obtained via an AL strategy based on the conditional marginals of the unlabeled samples, which encode the spatial information embedded in the DRF. The marginals are computed via the loopy belief propagation (LBP) method [25], [26], which is used in conjunction with labeled training samples in order to infer the class distributions. The main innovative contributions of this paper to the hyperspectral image classification literature can be summarized as follows.

1) First of all, we present an LBP-based method to estimate the conditional marginals for improved spectral–spatial hyperspectral image classification.
2) Second, by taking advantage from the aforementioned marginals, we apply AL algorithms in a more effective way as they can exploit all the spectral and the spatial information contained in the original hyperspectral data.

The remainder of this paper is organized as follows. Section II formulates the considered problem and its *maximum a posteriori* (MAP) solution. In Section III, we first estimate the marginals using the LBP algorithm. Then, we present several AL approaches which use both the spectral and spatial-contextual information. Section IV describes an experimental evaluation of the proposed approach, conducted using real hyperspectral data sets respectively collected by AVIRIS (over the Indian Pines region in Indiana) and by Reflective Optics System Imaging Spectrometer (ROSIS) (over the city of Pavia, Italy). Comparisons with state-of-the-art techniques for hyperspectral image classification are also reported. Finally, Section V concludes with some remarks and hints at plausible future research.

II. Considered Problem

First of all, we define the notation that will be adopted throughout this paper (Table I). Let $\mathcal{K} \equiv \{1, \ldots, K\}$ denote a set of K class labels; let $\mathcal{S} \equiv \{1, \ldots, n\}$ denote a set of integers indexing the n pixels of a hyperspectral image; let $\mathbf{x} \equiv (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathbb{R}^d$ denote such hyperspectral image made up of d-dimensional feature vectors; let $\mathbf{y} \equiv (y_1, \ldots, y_n)$ denote an image of labels; let $\mathcal{D}_L \equiv \{(y_1, \mathbf{x}_1), \ldots, (y_L, \mathbf{x}_L)\}$ be a set of labeled samples.

With these definitions in place, we can now build the posterior density $p(\mathbf{y}|\mathbf{x})$ of the class labels $\mathbf{y}$ given the features $\mathbf{x}$, which is the engine for the class label inference. We follow a discriminative approach. That is, we model the distribution $p(\mathbf{y}|\mathbf{x})$ directly, instead of the joint distribution $p(\mathbf{y}, \mathbf{x})$, which quite often implies simplistic assumptions about the data generation mechanism. Furthermore, because the discriminative models are less complex than the corresponding generative counterparts, learning the former models yields often better results than learning the latter ones, namely, when the training data are limited.

We adopt the following DRF model [19] for our posterior density[1]:

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x}, \boldsymbol{\omega})} \exp\left(\sum_{i \in \mathcal{S}} \log\, p(y_i|\mathbf{x}_i, \boldsymbol{\omega}) + \sum_{(i,j) \in \mathcal{C}} \log\, p(y_i, y_j) \right) \quad (1)$$

where $Z(\mathbf{x}, \boldsymbol{\omega})$ is the partition function and $\mathcal{C}$ is a set of cliques.[2] In the discriminative model (1), the term $\log\, p(y_i|\mathbf{x}_i, \boldsymbol{\omega})$, named the association potential, models the likelihood of label y_i given the feature vector $\mathbf{x}_i$, and the term $\log\, p(y_i|\mathbf{x}_i, \boldsymbol{\omega})$, named interaction potential, encodes spatial-contextual information.

In the DRF model introduced in [19], the association and interaction potentials are allowed to depend on the complete image of features. Our DRF (1) has a simpler structure as the association potential at pixel i depends only on the feature $\mathbf{x}_i$ and the interaction potential does not depend at all on the feature vectors. In spite of this simplification, our DRF model is able to capture to a great extent the spectral and the spatial information present in the hyperspectral data.

[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.

[2] A clique is a set of labels which are neighbors of each other.

TABLE I
LIST OF ABBREVIATIONS USED IN THIS PAPER

AIS	Airborne imaging spectrometer
AVIRIS	Airborne visible infra-red imaging spectrometer
ROSIS	Reflective optics system imaging spectrometer
AL	Active learning
DRF	Discriminative random field
MLR	Multinomial logistic regression
MRF	Markov random field
MLL	Multi-level logistic
MAP	*Maximum a posteriori*
MPM	*Maximum a posteriori* marginal
RBF	Radial basis function
SVM	Support vector machine
LBP	Loopy belief propagation
LORSAL	Logistic regression via splitting and augmented Lagrangian
LORSAL-MLL	Spectral-spatial segmentation via LORSAL and MLL
LORSAL-AL	LORSAL algorithm with active learning
LORSAL-AL-MLL	Spectral-spatial segmentation via LORSAL-AL and MLL
MPM-LBP	Maximizer of the posterior marginal by LBP
MPM-LBP-AL	Maximizer of the posterior marginal by LBP with AL
RS	Random selection
MI	Mutual information
BT	Breaking ties
MBT	Modified breaking ties

A. MLR

In this paper, we model the probability $p(y_i|\mathbf{x}_i, \boldsymbol{\omega})$ with the MLR

$$p(y_i = k|\mathbf{x}_i, \boldsymbol{\omega}) = \frac{\exp\left(\boldsymbol{\omega}^{(k)^T}\mathbf{h}(\mathbf{x}_i)\right)}{\sum_{k=1}^{K}\exp\left(\boldsymbol{\omega}^{(k)^T}\mathbf{h}(\mathbf{x}_i)\right)} \tag{2}$$

where $\boldsymbol{\omega} = [\boldsymbol{\omega}^{(1)^T}, \ldots, \boldsymbol{\omega}^{(K-1)^T}]^T$. Since the density in (2) does not depend on translations of the regressors $\boldsymbol{\omega}^{(k)}$, we take $\boldsymbol{\omega}^{(K)} = \mathbf{0}$. $\mathbf{h}(\mathbf{x}) = [h_1(\mathbf{x}), \ldots, h_l(\mathbf{x})]^T$ is a vector of l fixed functions of the input, often termed *features*. In this paper, we use a Gaussian radial basis function (RBF) kernel given by $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma^2)$, which is widely used in hyperspectral image classification problems [27]. In order to control the machine complexity and its generalization capacity, we model $\boldsymbol{\omega}$ as a random vector with Laplacian density [21]

$$p(\boldsymbol{\omega}) \propto \exp\left(-\lambda\|\boldsymbol{\omega}\|_1\right) \tag{3}$$

where λ is the regularization parameter controlling the degree of sparsity of $\boldsymbol{\omega}$.

Based on the prior $p(\boldsymbol{\omega})$ and on the DRF (1), we may derive an expectation–maximization algorithm to infer the MAP estimate of $p(\boldsymbol{\omega})$. This approach is heavy from the computational point of view, owing to the complexity associated to the expectation step and to the computation of the partition function $Z(\mathbf{x}, \boldsymbol{\omega})$ (see [19] and [28] for parameter learning in DRFs). In order to mitigate the complexity associated to the computation of the MAP estimate of $\widehat{\boldsymbol{\omega}}$, we assume that the interaction potential in (1) is constant, and thus, from (1), we have $p(\mathbf{y}|\mathbf{x}) = \prod_i p(y_i|\boldsymbol{x}_i)$, i.e., the random variables given $\boldsymbol{x}_i$, for $i = 1, \ldots, n$, are treated as independent. We stress that this assumption is taken only to infer the vector $\boldsymbol{\omega}$. We are aware that the inferred vector $\widehat{\boldsymbol{\omega}}$ is suboptimal. However, as we show in Section IV, this approximation is good enough to obtain state-of-the-art results. Therefore, we estimate $\boldsymbol{\omega}$ by

$$\widehat{\boldsymbol{\omega}} = \arg\max_{\boldsymbol{\omega}} \ell(\boldsymbol{\omega}) + \log\, p(\boldsymbol{\omega}) \tag{4}$$

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

$$\ell(\boldsymbol{\omega}) \equiv \sum_{i=1}^{L} \log\, p(y_i = k|\mathbf{x}_i, \boldsymbol{\omega}). \tag{5}$$

Optimization problem (4), although convex, is difficult to solve because the term of $\ell(\boldsymbol{\omega})$ is nonquadratic and the term $\log\, p(\boldsymbol{\omega})$ is nonsmooth. The sparse MLR (SMLR) algorithm presented in [21] solves this problem with $O((d(K-1))^3)$ complexity, which is frequently unbearable when dealing with kernels and a large number of classes. The LORSAL algorithm introduced in [22] and [29] opens a door to handle these by replacing a difficult nonsmooth convex problem with a sequence of quadratic plus diagonal $l_2 - l_1$ problems, resulting in a practical complexity cost of $O(d^2(K-1))$. Compared with the figure of $O((d(K-1))^3)$ associated to the SMLR algorithm, the complexity reduction is $d(K-1)^2$.

B. MLL Spatial Prior

In this paper, we include spatial-contextual information in the classification process by adopting an isotropic MLL prior to model the image of class labels $\mathbf{y}$. This prior, which belongs to the MRF class, encourages piecewise smooth segmentations and thus promotes solutions in which adjacent pixels are likely to belong to the same class. The MLL prior is a generalization of the Ising model [30] and has been widely used in image segmentation problems [23]. It is given by the following expression:

$$p(\mathbf{y}) = \frac{1}{Z} e^{\mu \sum\limits_{(i,j)\in\mathcal{C}} \delta(y_i - y_j)} \tag{6}$$

where Z is a normalizing constant for the density, μ is a tunable parameter controlling the degree of smoothness, and $\delta(y)$ is the unit impulse function.[3] 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 μ controls the degree of smoothness.

C. MAP Labeling

The MAP estimate minimizes the Bayesian risk associated to the zero–one loss function. Suppose that we are given the estimates $\widehat{\boldsymbol{\omega}}$ and $\widehat{\mu}$ of $\boldsymbol{\omega}$ and μ, respectively, and we want to compute the MAP estimate of $\mathbf{y}$ given by

$$\widehat{\mathbf{y}} = \arg\min_{\mathbf{y}\in\mathcal{K}^n} \left\{ \sum_{i\in\mathcal{S}} -\log\, p(y_i|\mathbf{x}_i, \widehat{\boldsymbol{\omega}}) - \widehat{\mu} \sum_{(i,j)\in\mathcal{C}} \delta(y_i - y_j) \right\}. \tag{7}$$

Minimization of (7) 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 mint-cuts on suitable graphs [31].

[3] i.e., $\delta(0) = 1$ and $\delta(y) = 0$, for $y \neq 0$.

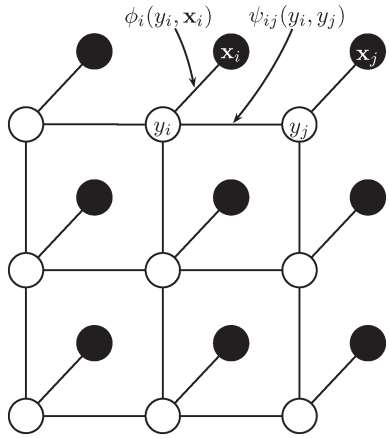

Fig. 1. Graphical example of a square lattice pairwise MRF.

This aspect has been thoroughly explored in the context of hyperspectral image classification in previous contributions [29].

III. Proposed Approach

A. MPM Labeling

An alternative to the MAP solution is the MAP marginal (MPM) solution, which minimizes the Bayesian risk associated to the sitewise zero-one loss function. As in the previous section, suppose that we are given the estimates $\widehat{\boldsymbol{\omega}}$ and $\widehat{\mu}$ of $\boldsymbol{\omega}$ and μ, respectively. The MPM estimate of label y_i is given by

$$\widehat{y}_i = \arg\max_{y_i} q(y_i|\mathbf{x}), \qquad i \in \mathcal{S} \tag{8}$$

where $q(y_i|\mathbf{x})$ is the marginal density of $p(\mathbf{y}|\mathbf{x})$ with respect to y_i. The computation of the marginal density of $p(\mathbf{y}|\mathbf{x})$ given by (1) is very difficult. In this paper, we use the LBP algorithm to estimate the MPM solution. LBP is an efficient inference approach to estimate Bayesian beliefs [25] in graphical models. The main idea of LBP is to introduce messages between hidden nodes in the MRF model as shown in Fig. 1, in which each node i represents a random variable or a hidden node, in our case, the class label y_i associated with each input feature vector $\mathbf{x}_i$. In the square lattice, $\psi_{ij}(y_i, y_j) = p(y_i, y_j)$ stands for the interaction potential that penalizes every dissimilar pair of neighboring labels, and $\phi_i(y_i, \mathbf{x}_i) = p(y_i|\mathbf{x}_i)$ represents the association potential of label y_i given evidence $\mathbf{x}_i$.

Fig. 2 provides a graphical interpretation in the form of a simple undirected network. At iteration t, the message sent from node i to any of its neighboring nodes $j \in \mathcal{N}(i)$ is given by

$$m_{ij}^t(y_j) = \frac{1}{Z} \sum_{y_i} \psi(y_i, y_j) \phi(y_i, \mathbf{x}_i) \prod_{k\in\mathcal{N}(i)\setminus\{j\}} m_{ki}^{t-1}(y_i) \tag{9}$$

where Z is a normalization constant. As shown in Fig. 2, at each iteration, each node sends messages to its neighbors, and the belief is estimated at each node by using all of the incoming messages. LBP is an iterative algorithm, where the propagation process iterates until convergence. The full joint posterior probability can be obtained from the product of marginals, or individual marginals can be used in an inference problem. If

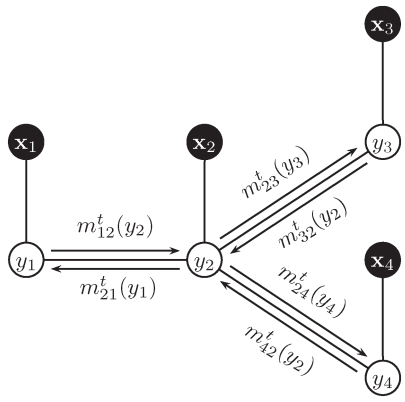

Fig. 2. Graphical representation of the message passing in LBP at iteration t.

we assume that $b_i^t(y_i)$ denotes the belief of node i at iteration t, then $b_i^t(y_i)$ is given by

$$b_i^t(y_i=k)=q(y_i=k|\mathbf{x}) \equiv \phi(y_i=k) \prod_{j\in\mathcal{N}(i)} m_{ji}^t(y_i=k). \quad (10)$$

Finally, we can estimate the final solution by means of a so-called *maximizer of the posterior marginal* for node i, which is given by

$$\widehat{y_i} \equiv \arg\max_{y_i} q(y_i|\mathbf{x}) \equiv \arg\max_{y_i} b_i^t(y_i). \quad (11)$$

Algorithm 1 presents the pseudocode for a supervised classification algorithm which takes advantage of the MLR, MLL, and LBP concepts to produce an MPM solution. In the required parameters of Algorithm 1, LBPiters denotes the number of iterations used in the LBP algorithm. In line 2 of Algorithm 1, the MLR regressors are learnt using the LORSAL algorithm [22]. In line 3 of Algorithm 1, $\mathbf{p_1}$ collects the MLR probabilities for the whole image, as indicated in (2). In line 4 of Algorithm 1, $\mathbf{p_2}$ collects the MLL priors as indicated in (6). Finally, in line 5 of Algorithm 1, the LBP algorithm is applied, which computes the marginal inference for the whole image. In [26], evidence is given that the LBP converges to a local stationary point of the Bethe approximation to the free energy. In our experiments, we empirically observed that LBP usually converges very fast, in practice, within less than ten iterations.

Algorithm 1 Maximizer of the posterior marginal by LBP (MPM-LBP)

Require: $\mathbf{x}$, $\mathcal{D}_L$, LBPiters, μ

1: $\widehat{\boldsymbol{\omega}} = \text{LORSAL}(\mathcal{D}_L)$ ($*$ LORSAL infers the sparse vectors $\boldsymbol{\omega}$ for MLR, see problem (4)$*$)
2: $\widehat{\mathbf{p}}_1 := \mathbf{p}_1(\mathbf{x}, \widehat{\boldsymbol{\omega}})$ ($*$ Function $\mathbf{p}_1$ collects the MLR probabilities, see (2) $*$)
3: $\widehat{\mathbf{p}}_2 := \mathbf{p}_2(\mu, \text{neighborhood})$ ($*$ Function $\mathbf{p}_2$ collects the MLL priors, see (6) $*$)
4: $\widehat{\mathbf{q}} := \text{LBP}(\widehat{\mathbf{p}}_1, \widehat{\mathbf{p}}_2, \text{LBPiters})$ ($*$ LBP estimates the marginals, see problem (10) $*$)
5: $\widehat{\mathbf{y}} := \arg\max_{\mathbf{y}} \widehat{\mathbf{q}}$

B. AL

The basic idea of AL is that of iteratively enlarging the training set by requesting an expert to label feature vectors from a set of unlabeled feature vectors. A relevant question is, of course, which samples should be chosen. This issue has been widely studied in the remote sensing literature [10]–[17], [29]. Most available approaches for AL take full advantage from the class posterior probabilities $p(y_i|\mathbf{x}_i)$. However, these methods usually learn from classifiers which only consider the spectral information, i.e., the given feature vector related with the class label. This is mainly because it is difficult to encode the spatial information in class posterior probability distributions. Furthermore, the posterior marginals $q(y_i|\mathbf{x})$ are very difficult to estimate. This is because computing the marginals of the joint distribution (1) represents a complex problem, which is very difficult to solve when the underlying graph contains cycles (loops). As a result, it is not easy to integrate spatial information using probabilities. To address this issue, in this paper, we resort to LBP to approximate such marginals. By taking advantage of the MPM estimates discussed in Section III-A, our introspection is that spatial information can also play a very important role in AL. In the following, we describe a new spectral–spatial strategy with AL which is based on the posterior marginal distributions $q(y_i|\mathbf{x})$ estimated by the LBP. With this strategy, we can implicitly include spatial information into existing AL approaches. The proposed spectral–spatial strategy will be shown in this paper to introduce improvements with regard to the most common case in which AL only considers the spectral information contained in the input data.

Algorithm 2 presents the pseudocode for the MPM-LBP algorithm with AL (MPM-LBP-AL hereinafter). Here, $\mathcal{D}_{L_i} \equiv \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_{L_i}, y_{L_i})\}$ denotes the initial labeled training set, and L_i is the number of samples in $\mathcal{D}_{L_i}$. Similarly, $\mathcal{D}_{L_u} \equiv \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_{L_u}, y_{L_u})\}$ is the new labeled set, and L_u is the number of new samples selected at each iteration during the AL process. Let $\mathcal{D}_U \equiv \{(\mathbf{x}_{L+1}, y_{L+1}), \ldots, (\mathbf{x}_{L+U}, y_{L+U})\}$ denote the candidate set for the AL set. Therefore, $\mathcal{D}_{L_u}$ is a subset of $\mathcal{D}_U$. As shown by Algorithm 2, in line 2, the marginal inferences are estimated by the MPM-LBP algorithm. In line 3, AL is performed in order to label new samples based on the posterior density $p(\mathbf{y}|\mathbf{x})$. Following our previous work [29], we consider four different sampling schemes for the AL step:

1) random selection (RS), where the new samples are randomly selected from the candidate set;
2) a mutual information (MI)-based criterion [32], [33], which aims at finding the samples maximizing the MI between the MLR regressors and the class labels;
3) a breaking ties (BT) algorithm [34], which aims at finding the samples minimizing the distance between the first two most probable classes;
4) a modified BT (MBT) scheme [29], which aims at finding samples maximizing the probability of the large class for each individual class.

It should be noted that, while there may be no diversity in the selection of unlabeled samples using RS and MI, the use of BT and MBT approaches includes diversity in the selection of unlabeled samples based on parameter L_u, which controls the

number of unlabeled samples included per iteration. As shown in our work [29], these methods have been traditionally applied using spectral information alone. In this paper, we will evaluate these methods using both spectral and spatial information based on the proposed framework. In line 4 of algorithm 2, the labeled training set is augmented by adding the set of newly selected labels. In line 5, the newly selected set of labels is removed from the candidate set in order to avoid overlapping effects.

Algorithm 2 Maximizer of the posterior marginal by LBP with AL (MPM-LBP-AL)

Require: $\mathbf{x}$, $\mathcal{D}_{L_i}$, $\mathcal{D}_L$, $\mathcal{D}_U$, BPiters, μ, L_u

1: **repeat**
2: $\widehat{\mathbf{q}} := \text{MPM-LBP}(\mathcal{D}_L, \mathbf{x}, \mu, \text{BPiters})$
3: $\mathcal{D}_{L_u} := \text{AL}(\widehat{\mathbf{p}}, \mathcal{D}_U)$
4: $\mathcal{D}_L := \mathcal{D}_L + \mathcal{D}_{L_u}$
5: $\mathcal{D}_U := \mathcal{D}_U - \mathcal{D}_{L_u}$
6: **until** some stopping criterion is met

To conclude this section, we emphasize that our proposed approach should not be regarded as a combination of many classifiers but as a fully unified algorithm with different steps, which has been carefully proposed bearing in mind the probabilistic nature of the techniques adopted, which use LBP to include spatial information as an innovative contribution in this paper. As a result, the proposed approach should be regarded as a unified framework for spectral–spatial classification of hyperspectral data in which AL techniques exhibit new potential, as they can exploit spatial information in conjunction with spectral information. In this regard, all parts of the proposed framework are considered equally important. In the following section, we experimentally evaluate the proposed approach using real hyperspectral data sets collected by different instruments.

IV. Experimental Results

In this section, we evaluate the proposed method by real hyperspectral data sets. In all of our experiments, we apply the Gaussian RBF kernel to a normalized version of the considered hyperspectral data set.[4] In our experiments, we fix the sparsity parameter to $\lambda = 0.001$ and the smoothness parameter to $\mu = 2$. A full discussion on the impact of these parameters on the final results is already given in our previous work [29]. Here, we concluded that, although the adopted parameter settings might be suboptimal, they empirically lead to very good results. We should also emphasize that, in our previous work [29], we observed that the use of cross-validation techniques yielded only marginal gains in the classification accuracies. Specifically, we have observed that the LORSAL algorithm is robust with respect to parameter λ and the LORSAL-MLL algorithm is robust with respect to parameter μ. As a result, we have decided to use fixed parameter settings in this paper in order to concentrate more on the proposed framework and less on its parameter tuning and also to illustrate that our proposed approach is not very sensitive to such parameter settings. Furthermore, we emphasize that, in this paper, we use first-order neighborhood connectivity in all experiments as we empirically found out that no significant gains in classification accuracy could be used by adopting higher order neighborhood systems which, in turn, require higher computational complexity.

Our main focus in experiments is to evaluate the proposed spectral–spatial-based methodology for solving MPM inferences in comparison with other approaches, such as those implemented by the MLR-based LORSAL algorithm [22], or the LORSAL-MLL [29], which implements the MAP strategy in (7), under the same experimental setup. Simultaneously, we also address the spectral-based method, such as LORSAL-AL and LORSAL-AL-MLL [29], for comparison. Finally, we also include other widely used classifiers in our comparison, such as the SVM with and without spatial preprocessing using extended morphological profiles (EMPs) and a spectral-spatial classifier based on the morphological watershed transform.

Before describing our experiments, we first reiterate the notations adopted. In the following, we assume that $\mathcal{D}_{L_i}$ denotes the initial labeled set, which is a subset of the available training set and that L_i denotes the number of samples (recall that L denotes the total number of labeled samples). Let $\mathcal{D}_U$ be the remaining samples in the ground-truth image. In all experiments, the initial set of labeled samples D_{L_i} is obtained by RS from the available ground-truth image, while the remaining samples $\mathcal{D}_U$ are used for testing purposes. These samples also served as the candidate pool for AL. Similarly, L_u denotes the number of newly selected labeled samples per iteration in the AL process, and D_{L_u} is a subset of $\mathcal{D}_U$. It is important to emphasize that, when D_{L_u} is selected at each iteration, we remove it from $\mathcal{D}_U$, i.e., $\mathcal{D}_U \equiv \mathcal{D}_U - \mathcal{D}_{L_u}$. In practice, we assume that the initial training samples for each class are uniformly distributed. It should be finally noted that, in all cases, the reported figures of overall accuracy (OA), average accuracy (AA), κ statistic, and class individual accuracies are obtained by averaging the results obtained after conducting ten independent Monte Carlo runs with respect to $\mathcal{D}_{L_i}$.

The remainder of this section is organized as follows. In Section IV-A, we introduce the data sets used for evaluation, which comprise the AVIRIS Indian Pines and ROSIS University of Pavia data sets (two widely used benchmarks for hyperspectral image classification). Section IV-B describes the experiments with the AVIRIS Indian Pines data set. Finally, Section IV-C conducts experiments using the ROSIS University of Pavia data set.

A. Hyperspectral Data Sets

Two hyperspectral data sets collected by two different instruments are used in our experiments.

1) The first hyperspectral image used in experiments was collected by the AVIRIS sensor over the Indian Pines region in Northwestern Indiana in 1992. This scene, with a size of 145 lines by 145 samples, was acquired over a mixed agricultural/forest area, early in the growing season. The scene comprises 220 spectral channels in the

[4]The normalization is simply given by $\mathbf{x}_i := \mathbf{x}_i/(\sqrt{\sum \|\mathbf{x}_i\|^2})$, for $i = 1, \ldots, n$, where $\mathbf{x}_i$ is a spectral vector.

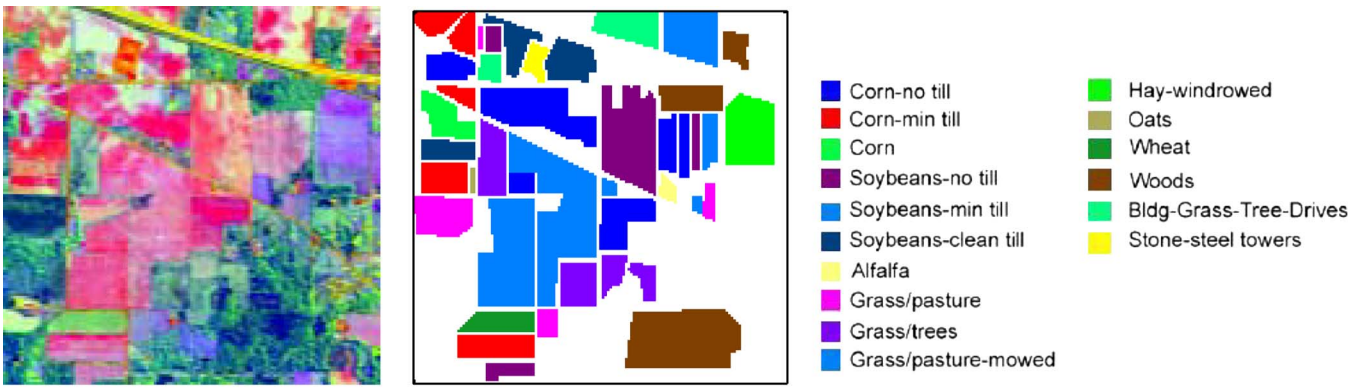

Fig. 3. (a) False color composition of the AVIRIS Indian Pines scene. (b) (Right) Ground-truth map containing 16 mutually exclusive land-cover classes.

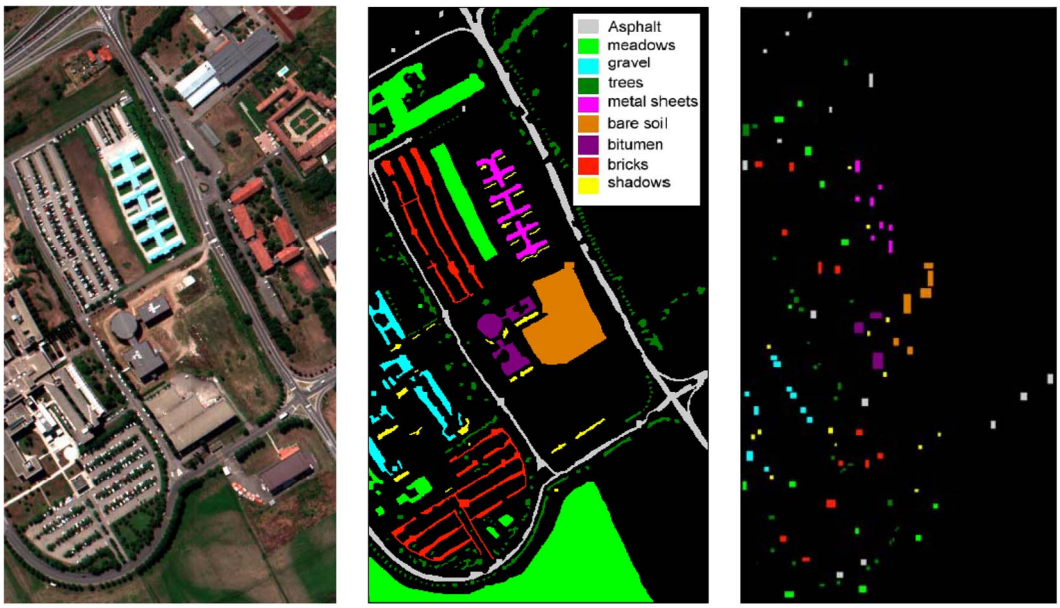

Fig. 4. (a) False color composition of the ROSIS University of Pavia scene. (b) Reference map containing nine mutually exclusive land-cover classes. (c) Training set used in experiments.

wavelength range from 0.4 to 2.5 μm, nominal spectral resolution of 10 nm, moderate spatial resolution of 20 m by pixel, and 16-b radiometric resolution. After an initial screening, several spectral bands were removed from the data set due to noise and water absorption phenomena, leaving a total of 200 radiance channels to be used in the experiments. For illustrative purposes, Fig. 3(a) shows a false color composition of the AVIRIS Indian Pines scene, while Fig. 3(b) shows the ground-truth map available for the scene, displayed in the form of a class assignment for each labeled pixel, with 16 mutually exclusive ground-truth classes, in total, 10 366 samples. These data, including ground-truth information, are available online,[5] a fact which has made this scene a widely used benchmark for testing the accuracy of hyperspectral data classification algorithms. This scene constitutes a very challenging classification 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.

2) The second hyperspectral data set was collected by the ROSIS optical sensor over the urban area of the University of Pavia, Italy. The flight was operated by the Deutschen Zentrum for Luftund Raumfahrt (DLR, the German Aerospace Agency) in the framework of the HySens project, managed and sponsored by the European Union. 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. 4(a) shows a false color composite of the image, while Fig. 4(b) shows nine reference classes of interest, which comprise urban features, as well as soil and vegetation features. Out of the available reference pixels, 3921 samples were used for training [see Fig. 4(c)], and 42 776 samples were used for testing. Some of the training samples in Fig. 4(c) are also included in the ground-truth classes in Fig. 4(b); hence, the total number of different labeled samples available for the ROSIS University of Pavia scene is 43 923. These training

[5]Available online: http://dynamo.ecn.purdue.edu/biehl/MultiSpec

TABLE II
OA, AA, INDIVIDUAL CLASS ACCURACIES (IN PERCENT), AND κ STATISTIC OBTAINED FOR DIFFERENT CLASSIFICATION METHODS WHEN APPLIED TO THE AVIRIS INDIAN PINES HYPERSPECTRAL DATA SET BY USING 10% OF THE GROUND-TRUTH DATA AS TRAINING SAMPLES

Class	Samples	Methods			
		SVM	LORSAL	LORSAL-MLL	MPM-LBP
Alfalfa	54	94.48	84.91	91.74	96.97
Corn-no till	1434	71.34	79.05	89.91	93.24
Corn-min till	834	68.00	69.43	87.55	90.68
Corn	234	85.23	86.38	97.82	99.59
Grass/pasture	497	73.19	94.02	95.74	96.66
Grass/tree	747	96.94	96.78	99.30	99.52
Grass/pasture-mowed	26	77.77	91.40	96.92	100
Hay-windrowed	489	85.89	99.04	99.38	99.52
Oats	20	73.74	96.00	100	100
Soybeans-no till	968	86.31	79.39	92.20	93.67
Soybeans-min till	2468	87.03	73.84	90.15	93.21
Soybeans-clean till	614	92.71	87.54	95.99	97.50
Wheat	212	94.27	99.20	99.40	99.40
Woods	1294	97.42	92.33	94.04	94.48
Bldg-grass-tree-drives	380	99.13	78.59	93.54	97.39
Stone-steel towers	95	90.29	93.11	98.56	98.84
OA		80.56	82.60	92.72	94.76
AA		77.81	87.56	95.14	96.92
κ		85.86	80.14	91.66	93.99

and test sets are widely used in the hyperspectral image classification community and provided by the University of Pavia, who conducted the ground-truth data collection and labeled sample generation for this particular scene.

B. Experiments With AVIRIS Indian Pines Data Set

In this set of experiments, we first evaluated the classification accuracy of the proposed MPM-LBP algorithm. Table II shows the OA, AA, κ statistic, and individual class accuracies obtained using 1036 training samples, i.e., 10% of each class in the ground-truth image shown in Fig. 4(b). It is worth noting that, in this challenging classification scenario, the MPM-LBP algorithm achieved better performance than the other spectral–spatial algorithm, i.e., LORSAL-MLL [29], considered for comparison here. Furthermore, the MPM-LBP significantly improved algorithms considering the spectral information alone, such as the SVM [6] and LORSAL [22]. Overall, the experiments suggest that the proposed MPM-LBP is competitive with some of the best available spectral–spatial methods for hyperspectral image classification.

In a second experiment, we evaluated the proposed MPM-LBP-AL algorithm, which performs spectral–spatial-based AL, and further compared it with other spectral-based AL strategies. Fig. 5 plots the obtained OA classification results for four different AL approaches: RS, MI, BT, and MBT, as functions of the number of labeled training samples for the AVIRIS Indian Pines scene. In all cases, we started with an initial training set of 80 samples (only five per class), i.e., $L_i = 80$ and $L_u = 10$ (ten new samples were included at each iteration). These algorithms are integrated in the learning strategies considered in the previous section, i.e., LORSAL (resulting in a strategy called LORSAL-AL) and LORSAL-MLL (resulting in a strategy called LORSAL-AL-MLL). In those cases, the AL methods only use spectral information. Finally, we also integrated the AL approaches into the MPM-LBP, resulting in a spectral–spatial AL strategy called MPL-LBP-AL.

Several conclusions can be obtained from Fig. 5. First and foremost, the proposed MPM-LBP-AL algorithm clearly obtained the best results in all cases. This is due to the fact that the proposed approach takes advantage of both spatial and spectral information, as opposed to the other tested approaches

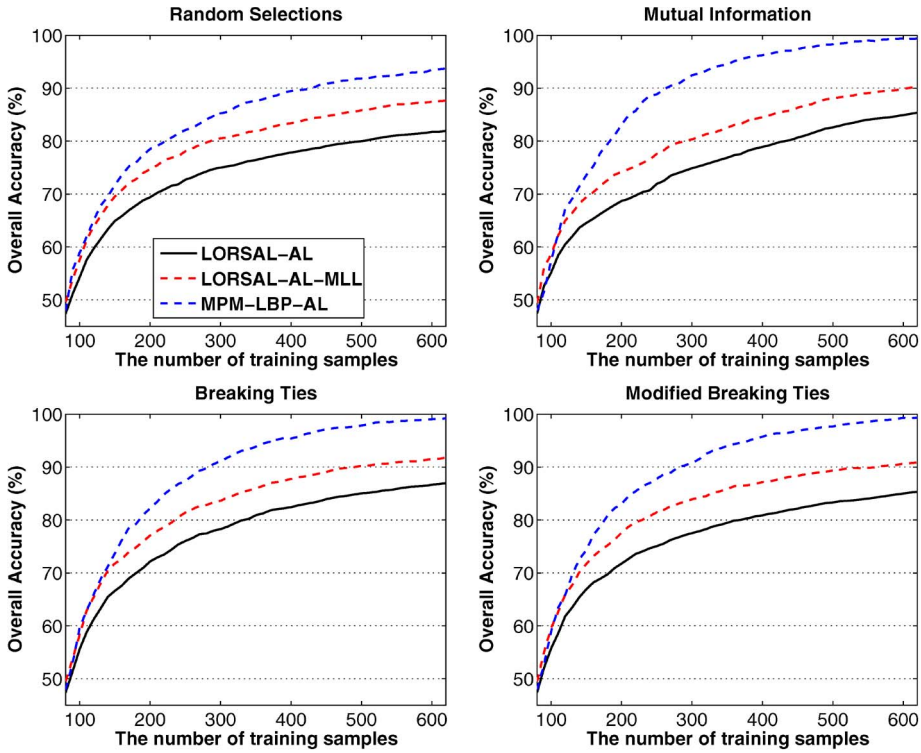

Fig. 5. OA classification results (as a function of the number of labeled training samples) achieved by the LORSAL-AL, LORSAL-AL-MLL, and MPM-LBP-AL for the AVIRIS Indian Pines scene. Here, an initial training set of 80 labeled samples (five per class) is used, i.e., $L_i = 80$, and ten new samples are included at each iteration, i.e., $L_u = 10$.

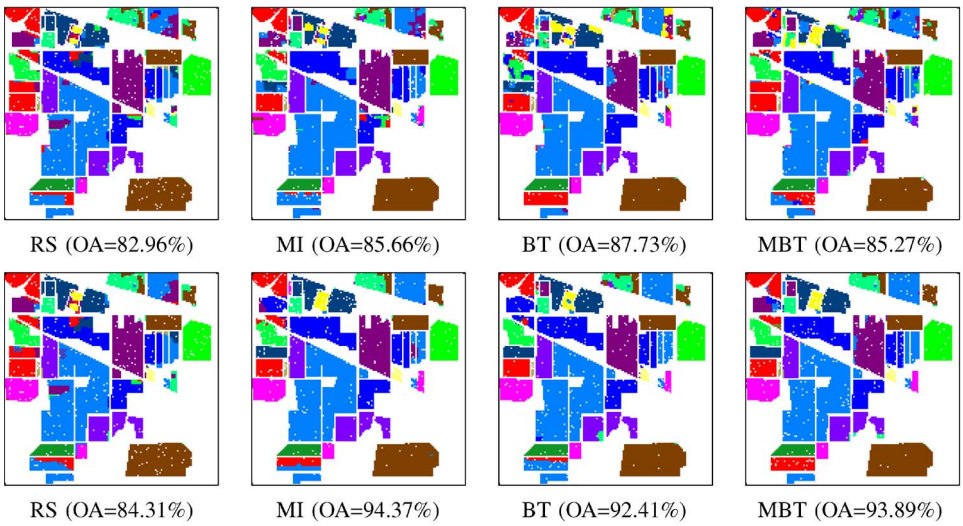

Fig. 6. Classification maps obtained from AVIRIS Indian Pines data set (the training samples selected are superimposed as white dots in the classification maps) with $L_i = 80$ samples and a total number of 280 samples for (top) LORSAL-AL-MLL and (bottom) MPM-LBP-AL.

which use spectral information alone. For example, with 600 labeled samples, all tested sampling methods converged to OAs higher than 99% when MPM-LBP-AL was used. However, for the LORSAL-AL-MLL algorithm in which the sampling is exclusively based on spectral information, the OAs converged to around 91%. It is also worth noting that the performance of the proposed MPM-LBP-AL depends on the number of labeled training samples. When L is very small, the proposed MPM-LBP-AL provides very similar results with regard to those obtained by LORSAL-MLL. As L increases, for example, to $L \geq 150$, the MPM-LBP-AL substantially increases its classification accuracy with regard to that of LORSAL-AL-MLL. Another interesting observation from Fig. 5 is that RS in Fig. 5(a) did not perform as effectively as the other tested approaches. This was expected, since the other sampling methods actively search for the most informative labeled samples. For a full discussion on the behavior of these sampling methods, we refer interested readers to [29].

In order to show the good performance of the proposed spectral–spatial AL approach, Fig. 6 shows the obtained classification maps along with the training samples selected by each AL method (superimposed as white dots on each classification map). It is noticeable that the proposed spectral-spatial AL strategy selected more samples in the difficult regions, such as

TABLE III
OA, AA, INDIVIDUAL CLASS ACCURACY (IN PERCENT), AND κ STATISTIC OBTAINED FOR DIFFERENT CLASSIFICATION METHODS WHEN APPLIED TO THE ROSIS PAVIA HYPERSPECTRAL DATA SET

Class	Number of samples		Methods					
	Training	Test	LORSAL[†]	LORSAL-MLL[‡]	MPM-LBP[♭]	SVM[♯]	EMP/SVM[♯]	Watershed[§]
Asphalt	548	6631	82.43	96.71	95.70	83.71	95.36	93.64
Bare soil	540	18649	69.08	72.36	73.27	92.25	63.72	97.35
Bitumen	392	2099	74.27	72.42	74.18	81.58	98.87	96.23
Bricks	524	3064	96.41	97.85	97.85	92.59	95.41	97.92
Gravel	265	1345	99.26	99.78	99.85	70.32	87.61	66.12
Meadows	532	5029	92.96	98.35	98.55	70.25	80.33	75.09
Metal sheets	375	1330	89.85	98.50	97.97	99.41	99.48	99.91
Shadows	514	3682	89.54	99.29	98.89	96.62	97.68	96.98
Trees	231	947	95.46	97.57	93.56	97.81	98.37	98.56
OA			80.11	85.57	85.78	80.99	85.22	85.42
AA			87.70	92.54	92.20	88.28	90.76	91.31
κ			75.09	81.80	82.05	76.16	80.86	81.30

[†] LORSAL algorithm implements an MLR-based solution [22];

[‡] LORSAL-MLL algorithm implements the MAP solution in (7) [29];

[♭] MPM-LBP algorithm is the proposed spectral-spatial method;

[♯] SVM results are taken from [6], which used EMPs for spectral-spatial characterization prior to SVM-based classification;

[§] Watershed results are taken from [36], 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.

the upper leftmost quadrant of the image which is dominated by highly mixed classes. The resulting classification maps are smoother than the MAP solutions. For example, the MAP solutions in the upper leftmost quadrant exhibit significant confusion between classes.

C. Experiments With ROSIS University of Pavia Data Set

In this experiment, we conduct an evaluation of the proposed MPM-LBP algorithm with regard to other state-of-the-art hyperspectral image classification and segmentation methods. Table III shows the OA, AA, κ, and individual class accuracies obtained in our algorithm comparison, which includes well-known spectral-based classification methods such as the SVM [6] and LORSAL [22]. Given the importance of considering spatial information in the analysis of this particular scene, we have also considered three spectral–spatial methods in our comparison: an SVM-based classifier trained with EMPs (designated in the table by EMP/SVM) [35], a segmentation method based on the watershed transform [36], and LORSAL-MLL [29]. The results reported in the table are respectively taken from [35] and [36], where exactly the same training and test sets were used to produce the results reported in Table III, thus allowing a fair intercomparison of methods. As shown by Table III, the proposed MPM-LBP provides results which are comparable to those obtained by other spectral–spatial methods. From Table III, it can be observed that all spectral–spatial-based algorithms (LORSAL-MLL, MPM-LBP, EPM/SVM, and watershed) exhibit good performance, indicating the importance of including spatial information in the analysis.

In a second experiment, we evaluated the proposed spectral–spatial-based AL approach (MPM-LBP-AL). We use all available samples [a total of 43 923 samples obtained after combining all the different labeled samples from the ground-truth image in Fig. 4(b) and from the fixed training set in Fig. 4(c)] as the candidate set for the AL process. We started with an initial training set made up of 90 samples, i.e., $L_i = 90$ and $L_u = 10$ (ten new samples were included at each iteration). Again, LORSAL-AL and LORSAL-AL-MLL are considered for comparative purposes. Fig. 7 shows the obtained results as a function of the number of labeled training samples. Several conclusions can be obtained from Fig. 7. First of all, the proposed MPM-LBP-AL algorithm clearly obtained the best results in all cases, which illustrates the advantage of spectral–spatial-based AL. Furthermore, the proposed spectral-spatial strategy is more robust for AL. For example, all MI, BT, and MBT sampling approaches outperformed RS by the proposed spectral-spatial strategy. However, the MI and MBT approaches obtained slightly worse results when only spectral information is considered for AL.

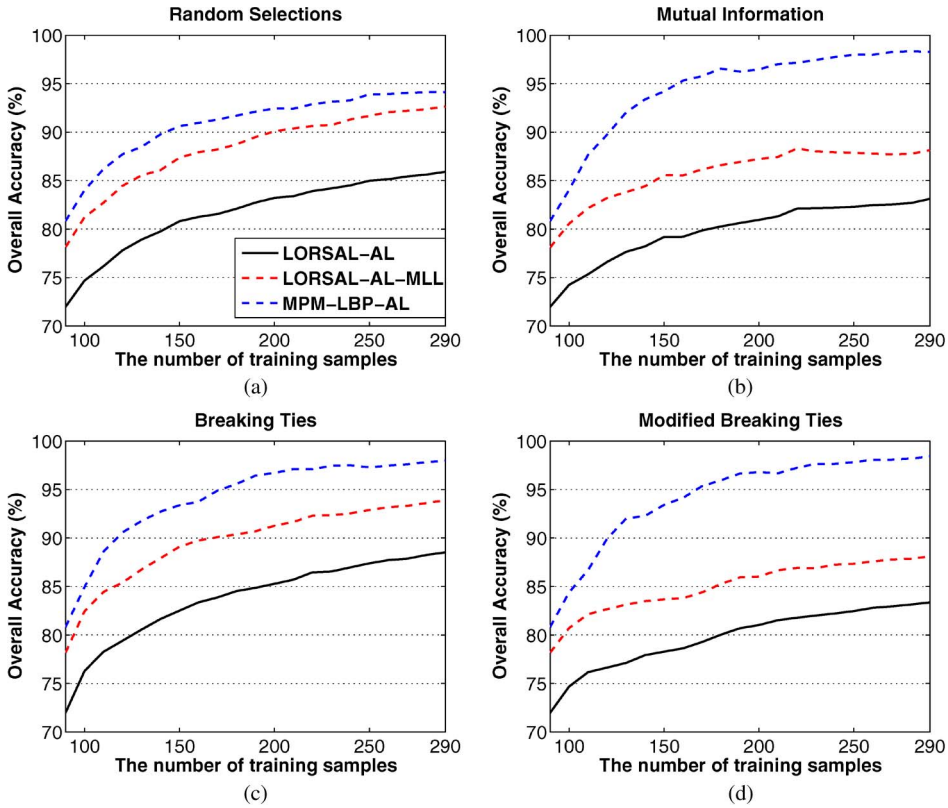

Fig. 7. OA classification results (as a function of the number of labeled training samples) achieved by the LORSAL-AL, LORSAL-AL-MLL, and MPM-LBP-AL for the ROSIS University of Pavia data set. Here, an initial training set of 90 labeled samples (ten per class) is used, i.e., $L_i = 90$, and ten new samples are included at each iteration, i.e., $L_u = 10$.

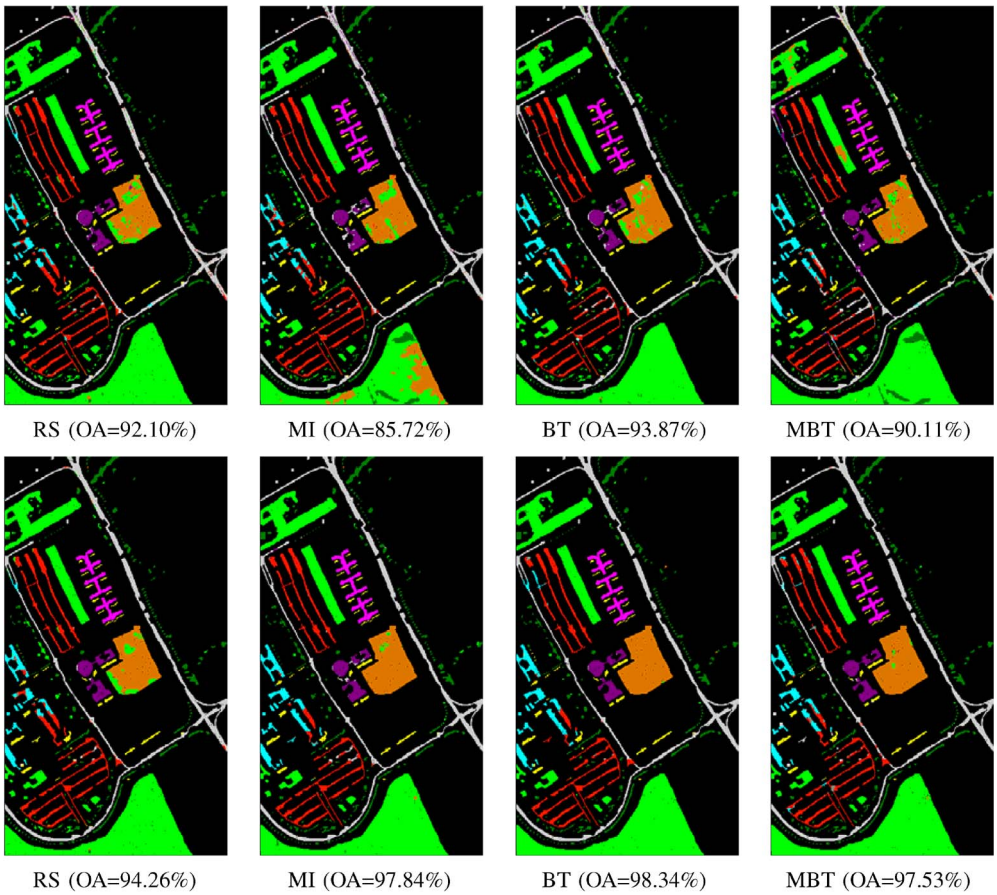

Fig. 8. Classification maps obtained from ROSIS University of Pavia data set along with (superimposed as black dots in the classification maps) the training set selected by each method with $L_i = 90$ samples and a final number of 290 samples for (top) LORSAL-AL-MLL and (bottom) MPM-LBP-AL.

Finally, Fig. 8 shows the classification maps (along with the training sets selected by each AL method, this time superimposed as black dots in the respective maps). It can be seen that the proposed spectral–spatial AL strategy outperforms the spectral-based sampling approaches as the resulting classification maps are smoother than the MAP solutions.

V. Conclusion and Future Research Lines

In this paper, we have proposed a new framework for spectral–spatial hyperspectral classification using belief propagation, in the context of which AL algorithms consider both spatial and spectral information simultaneously. According to our knowledge, this represents an innovative contribution in the AL literature as previous methods have been mainly based on the exploitation of spectral information alone. The parameters of the complete model adopted by our strategy are inferred via Bayesian inference using LBP tools. We highlight that the proposed integrated algorithm has been shown to be very well suited to problems with very few training samples available *a priori*. In the two considered analysis scenarios (AVIRIS Indian Pines and ROSIS University of Pavia data sets), the proposed method exhibited state-of-the-art performance with regard to other similar approaches. Although the results obtained are very encouraging, further experiments with additional scenes and comparison methods should be conducted. In the future, we will also develop computationally efficient implementations of the proposed approaches by resorting to parallel computer architectures such as commodity clusters or graphical processing units.

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 Pavia, Italy, along with the training and test sets, and V. Kolmogorov for the max flow/min-cut C++ code made available. Last but not least, the authors would also like to thank the Associate Editor who handled our paper and the two anonymous reviewers for providing truly outstanding comments and suggestions that significantly helped us to improve the technical quality and presentation of our paper.

References

[1] C.-I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. New York: Kluwer, 2003.

[2] A. F. Goetz, G. Vane, J. E. Solomon, and B. N. Rock, "Imaging spectrometry for Earth remote sensing," *Science*, vol. 228, no. 4704, pp. 1147–1153, 1985.

[3] 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.

[4] G. Shaw and D. Manolakis, "Signal processing for hyperspectral image exploitation," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 12–16, Jan. 2002.

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

[6] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. 110–122, Sep. 2009.

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

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

[9] D. Landgrebe, "Hyperspectral image data analysis as a high dimensional signal processing problems," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.

[10] S. Rajan, J. Ghosh, and M. M. Crawford, "An active learning approach to hyperspectral data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 4, pp. 1231–1242, Apr. 2008.

[11] D. Tuia, F. Ratle, F. Pacifici, M. F. Kanevski, and W. J. Emery, "Active learning methods for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2218–2232, Jul. 2009.

[12] G. Jun and J. Ghosh, "An efficient active learning algorithm with knowledge transfer for hyperspectral data analysis," in *Proc. Int. Geosci. Remote Sens. Symp.*, 2008, pp. I-52–I-55.

[13] L. Copa, D. Tuia, M. Volpi, and M. Kaneski, "Unbiased query-by-bagging active learning for VHR image classification," in *Proc. SPIE Eur. Remote Sens.*, 2010, pp. 78300K-1–78300K-8.

[14] B. Demir, C. Persello, and L. Bruzzone, "Batch-mode active-learning methods for the interactive classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 1014–1031, Mar. 2011.

[15] S. Patra and L. Bruzzone, "A fast cluster-assumption based active-learning technique for classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 5, pp. 1617–1626, May 2011.

[16] W. Di and M. M. Crawford, "Active learning via multi-view and local proximity co-regularization for hyperspectral image classification," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 618–628, Jun. 2011.

[17] S. Patra and L. Bruzzone, "A batch-mode active learning technique based on multiple uncertainty for SVM classifier," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 3, pp. 497–501, May 2012.

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

[19] S. Kumar and M. Hebert, "Discriminative random fields," *Int. J. Comput. Vis.*, vol. 68, no. 1, pp. 179–201, Jun. 2006.

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

[21] 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.

[22] J. M. Bioucas-Dias and M. Figueiredo, "Logistic regression via variable splitting and augmented Lagrangian tools," Inst. Superior Técnico, TU, Lisbon, Portugal, 2009, Tech. Rep.

[23] S. Z. Li, *Markov Random Field Modeling in Computer Vision*. London, U.K.: Springer-Verlag, 1995.

[24] 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.

[25] J. Yedidia, W. Freeman, and Y. Weiss, "Understanding belief propagation and its generalizations," in *Proc. Int. Joint Conf. Artif. Intell.*, Seattle, WA, 2001.

[26] J. S. Yedidia, W. T. 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.

[27] 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.

[28] S. Kumar and M. Hebert, "Discriminative random fields: A discriminative framework for contextual interaction in classification," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, 2003, pp. 1150–1157.

[29] 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.

[30] 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.

[31] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.

[32] D. Mackay, "Information-based objective functions for active data selection," *Neural Comput.*, vol. 4, no. 4, pp. 590–604, Jul. 1992.
[33] B. Krishnapuram, D. Williams, Y. Xue, A. Hartemink, L. Carin, and M. Figueiredo, "On semi-supervised classification," in *Proc. 18th Annu. Conf. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, 2004, pp. 721–728.
[34] T. Luo, K. Kramer, D. B. Goldgof, S. Samson, A. Remsen, T. Hopkins, and D. Cohn, "Active learning to recognize multiple types of plankton," *J. Mach. Learn. Res.*, vol. 6, pp. 589–613, Dec. 2005.
[35] 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.
[36] 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.

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 and the Instituto de Telecomunicações, 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. She is currently with the Hyperspectral Computing Laboratory research group coordinated by Prof. Antonio Plaza at the Department of Technology of Computers and Communications, University of Extremadura, Mérida, 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.

José M. Bioucas-Dias (S'87–M'95) received the B.S. degree in electrical engineering and the M.Sc., Ph.D., and Agregado degrees in electrical and computer engineering from the Universidade Técnica de Lisboa, Lisboa, Portugal, in 1985, 1991, 1995, and 2007, respectively.

Since 1995, he has been with the Departamento de Engenharia Electrotécnica e de Computadores, Instituto Superior Tecnico, Universidade Técnica de Lisboa, where 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 has been involved in several national and international research projects and networks, including the Marie Curie Actions Hyperspectral Imaging Network 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 TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING). He has been a member of program/technical committees of several international conferences. He was the General Cochair of the Third IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (2011).

Antonio Plaza (M'05–SM'07) received the M.S. and Ph.D. degrees in computer engineering from the University of Extremadura, Mérida, 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 Airborne Visible Infrared Imaging Spectrometer 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, where he is the Head of the Hyperspectral Computing Laboratory. He was the Coordinator of the Hyperspectral Imaging Network, 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 two special issues on remotely sensed hyperspectral imaging for different journals, including 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 also guest edited an IEEE Special Issue of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (for which he has served as an Associate Editor on hyperspectral image analysis and signal processing since 2007) and an IEEE Special Issue of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS 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 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. He is currently serving as the Director of Education Activities for the IEEE Geoscience and Remote Sensing Society.