Supervised Spectral–Spatial Hyperspectral Image Classification With Weighted Markov Random Fields

Le Sun, *Student Member, IEEE*, Zebin Wu, *Member, IEEE*, Jianjun Liu, Liang Xiao, and Zhihui Wei

***Abstract*—This paper presents a new approach for hyperspectral image classification exploiting spectral–spatial information. Under the maximum *a posteriori* framework, we propose a supervised classification model which includes a spectral data fidelity term and a spatially adaptive Markov random field (MRF) prior in the hidden field. The data fidelity term adopted in this paper is learned from the sparse multinomial logistic regression (SMLR) classifier, while the spatially adaptive MRF prior is modeled by a spatially adaptive total variation (SpATV) regularization to enforce a spatially smooth classifier. To further improve the classification accuracy, the true labels of training samples are fixed as an additional constraint in the proposed model. Thus, our model takes full advantage of exploiting the spatial and contextual information present in the hyperspectral image. An efficient hyperspectral image classification algorithm, named SMLR-SpATV, is then developed to solve the final proposed model using the alternating direction method of multipliers. Experimental results on real hyperspectral data sets demonstrate that the proposed approach outperforms many state-of-the-art methods in terms of the overall accuracy, average accuracy, and kappa (k) statistic.**

***Index Terms*—alternating direction method of multipliers (ADMM), hyperspectral classification (HC), sparse multinomial logistic regression (SMLR), spatially adaptive TV constraint.**

I. Introduction

HYPERSPECTRAL remote sensors capture digital images in hundreds of continuous narrow (about 2–10 nm) spectral bands spanning the visible to infrared spectrum (400–2500 nm). Pixels in hyperspectral imagery (HSI) are represented by vectors whose entries correspond to the spectral bands. Different materials usually reflect electromagnetic energy differently at specific wavelengths. Moreover, this allows the characterization, identification, and classification of the land covers with improved accuracy and robustness. Among all of these aforementioned applications, classification is one of the most popular ones in the analysis of hyperspectral image.

Recently, many techniques have been developed for HSI classification. Among these methods, discriminative approaches in machine learning fields have shown success in hyperspectral classification (HC) due to their advantages in dealing with high dimensionality and small-size training samples. For instance, multinomial logistic regression (MLR) based techniques are among the state-of-the-art discriminative techniques, as well as support vector machines (SVMs). As the ability of dealing with small-size training samples, these kinds of methods have been proved to be robust against the well-known Hughes effect [1], [2]. SVMs have shown remarkable success for supervised HSI classification [3]–[6]. In turn, MLR-based [7] approaches have the ability to learn the class distributions themselves in a Bayesian learning framework [8], thus supplying a degree of plausibility for such classes. Various sparse MLR (SMLR) methods are investigated in the literature [9]–[11]. More recently, the logistic regression via variable splitting and augmented Lagrangian (LORSAL) algorithm [10], which is able to learn a sparse regressor with Laplace prior distribution of SMLR, has become a new approach to deal with larger data sets and number of classes efficiently. Moreover, these ideas have been successfully applied to HSI classification [12]. For instance, a new semisupervised learning algorithm was proposed to deal with HSI classification [13], in which a soft SMLR model was learned from both the labeled and unlabeled samples. Due to its practical cost of computation for dealing with the very high dimension data, SMLR with LORSAL has been validated as one of the best approaches among all state-of-the-art techniques for pixel-only HC [11], [25], [26], [29].

However, the SMLR approaches only make use of the spectral information, which will probably lead to some outliers or classification errors. To cope with this problem, spectral–spatial classification techniques have recently drawn considerable attention. In fact, the materials of a certain class in HSI usually gather together. For a given pixel, if it belongs to one class, its neighboring pixels have high probability of belonging to the same class. Moreover, some structures (e.g., edges) can also be extracted from the spatial information. Therefore, the combination of spectral information and spatial context can improve the result of HSI classification. Many techniques have been proposed to impose the spatial information, such as

Manuscript received May 30, 2013; revised November 13, 2013, March 24, 2014, and July 15, 2014; accepted July 24, 2014. This work was supported by the National Natural Science Foundation of China under Grants 61101194 and 61071146, the Jiangsu Provincial Natural Science Foundation of China under Grant BK2011701, the Research Fund for the Doctoral Program of Higher Education of China under Grants 20113219120024 and 20123219120043, the Project of China Geological Survey under Grant 1212011120227, the Jiangsu Planned Projects for Postdoctoral Research Funds under Grant 0901008B, and the National Scientific Equipment Developing Project of China under Grant 2012YQ050250.

L. Sun, J. Liu, L. Xiao, and Z. Wei are with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: sunlecncom@163.com).

Z. Wu is with Jiangsu Key Laboratory of Spectral Imaging and Intelligent Sensing, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: zebin.wu@gmail.com).

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.2014.2344442

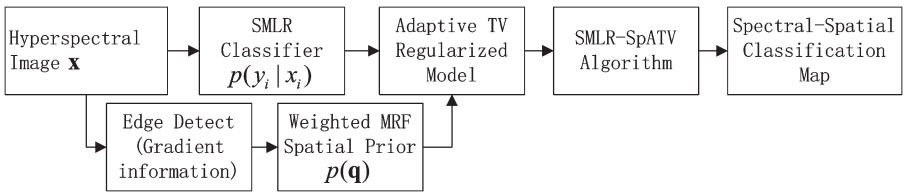


Fig. 1. Flowchart of our proposed SMLR-SpATV classification scheme.

morphological filters (e.g., [5]), partitional clustering (e.g., [14] and [15]), composite kernels (CKs) (e.g., [16]), graph kernels (e.g., [17] and [18]), joint sparse representation (e.g., [19] and [20]), and Markov random fields (MRFs) (e.g., [21] and [25]).

As a general image statistical modeling tool, MRFs are effective in incorporating the spatial information under Bayesian inferring framework and have shown good performance in HSI classification. For instance, in [21], the result of probabilistic SVM was regularized by an MRF. In [22], SVMs and MRFs were combined in an integrated framework for contextual HSI classification. In [23], an adaptive MRF approach was proposed for HSI classification, and a relative homogeneity index for each pixel was used to determine the weighting coefficient for the MRF classification. In [24] and [25], the framework of combining (subspace) SMLR with a multilevel logistic Markov–Gibbs MRF was proposed to classify the HSI. In [26], the HC results were obtained by maximizing the marginal probability of the posterior distribution using the loopy belief propagation method, where the posterior class probability was modeled as an MLR classifier and an MRF. Some of the other HSI classification techniques that take MRFs into consideration can also be found in [27] and [28]. Despite the success of the aforementioned MRF-based techniques in HSI classification, there is still much room for MRF-based approaches to improve the accuracy by further exploiting the spatial structure (e.g., edges) and contextual information. Moreover, these MRF-based spectral–spatial methods only focus on the spatial prior on the discrete-valued labels, thus leading to a difficult discrete optimization problem [42].

In [41], the idea of using a hidden set of real-valued fields opens a new door to sidestep the discrete nature of the image segmentation problem. Later, the hidden-field-based segmentation algorithms were further developed in [42] and [43] by using Gaussian field prior and wavelet-based priors, respectively. Inspired by these ideas, we propose a new framework (see Fig. 1) for hyperspectral image classification using spectral–spatial information under the real-valued hidden field. First, the spatial information is imposed by modeling the potential function associated with a weighted MRF as a spatially adaptive vector TV function, which has been proved to be effective for imposing spatial coherence in image processing [44], [45]. The vector TV function defined on the real-valued hidden marginal probability of the posterior distribution would enforce a piecewise smooth result, and the weights are calculated by the gradients of the original hyperspectral image to impose the spatial structure information. Second, the proposed maximum *a posteriori* (MAP) model is transformed into a spatially adaptive TV regularization model, which is solved by alternating direction method of multipliers (ADMM).

Different from the method proposed in [42] and [43], our method models the spatial prior on the hidden marginal probability by which the labels are totally determined using an MRF prior with an adaptive-TV-induced potential function, while the method proposed in [42] and [43] models the spatial prior on the hidden image/field to which the labels are probabilistically related using Gaussian field or wavelet-based prior. Moreover, our method is constrained (such as nonnegative constraint, sum-to-one constraint, and training label fixed constraint), while the method proposed in [42] and [43] is totally unconstrained. The most important one is that, by modeling the spatial prior on the hidden marginal probability, our model is finally transformed into an adaptive TV regularized variation problem and can be easily solved by ADMM, while the model proposed in [42] and [43] is approximately solved by the expectation-maximization algorithm [46].

The remainder of this paper is organized as follows. Section II formulates the HC problem in Bayesian framework using SMLR and MRFs to get its MAP solution and outlines the framework of our proposed method. Section III describes the proposed model imposing spatial context on the hidden marginal probability of the posterior distribution instead of the discrete-valued labels and presents the algorithm to solve it. Section IV reports the experimental results on the real hyperspectral data sets in comparison with other state-of-the-art supervised classifiers. Section V concludes with some remarks and outlines future work.

II. Problem Formulation

Let $\mathcal{S} \equiv \{1, 2, \ldots, N\}$ denote integer indexes of the N pixels in a hyperspectral image; let $\mathbf{x} \equiv (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N) \in R^{L \times N}$ be a hyperspectral image with N samples and L features (bands), where each $\boldsymbol{x}_i$ is an L-dimensional vector; and let $\mathbf{y} \equiv (y_1, y_2, \ldots, y_N) \in \mathcal{K}^N$ be an image of class labels, where $\mathcal{K} \equiv \{1, \ldots, K\}$ denotes a set of K class labels and each $y_i = [y_i^{(1)}, y_i^{(2)}, \ldots, y_i^{(K)}]$ is a 1-of-K encoding of the classes ($y_i^{(j)} \in \{0, 1\}$, for $j \in \mathcal{K}$ and K is the number of classes). With the aforementioned notations in place, the goal of HC is to assign a label $y_i \in \mathcal{K}$ to each pixel $i \in \mathcal{S}$ based on the observed hyperspectral signal $\boldsymbol{x}_i$, resulting in an image of class labels $\mathbf{y}$.

In this paper, the HC problem is studied in Bayesian framework. By adopting the discriminative random field model [12], [29], the MAP classification is then given by

$$\begin{aligned}\widehat{\mathbf{y}} &= \arg\max_{\mathbf{y} \in \mathcal{K}^N} \left(\log p(\mathbf{y}|\mathbf{x})\right) \\ &= \arg\max_{\mathbf{y} \in \mathcal{K}^N} \left\{ \sum_{i=1}^{N} \left(\log p(y_i|\boldsymbol{x}_i) - \log p(y_i)\right) + \log p(\mathbf{y}) \right\}. \end{aligned} \tag{1}$$

In machine learning field, the densities $p(y_i|\boldsymbol{x}_i)$ are usually learned by the probability classifiers. For example, in [24], the densities $p(y_i|\boldsymbol{x}_i)$ are modeled as MLRs, whose regressors are learned via the LORSAL algorithm [10], and in [21], they are modeled as probabilistic SVM classifiers. In contrast to MLRs and SVMs, in our previous work [30], they are obtained from the fully constrained sparse unmixing procedure. The densities $p(y_i)$, without loss of generality, are assumed to follow the uniform distribution (i.e., $p(y_i) = 1/K$ for each $i \in \mathcal{S}$). The density $p(\mathbf{y})$, in HC problem, is usually modeled to impose the spatial prior on the labels $\mathbf{y}$.

Specifically, the proposed approach in this paper can be viewed as an MRF regularized-SMLR classifier. Hence, we will first review the fundamentals of SMLR and then outline the framework of our proposed approach.

A. Review of SMLR

Supposing that $\boldsymbol{w}^{(k)}$ is the feature-weight vector corresponding to class k and according to the MLR model [9], the probability $p(y_i|\boldsymbol{x}_i, \boldsymbol{w})$ that a given sample $\boldsymbol{x}_i$ belongs to class k is formulated as follows:

$$p_i^{(k)} \equiv p\left(y_i^{(k)} = 1|\boldsymbol{x}_i, \boldsymbol{w}\right) = \frac{\exp\left(\boldsymbol{w}^{(k)^T} h(\boldsymbol{x}_i)\right)}{\sum_{j=1}^{K} \exp\left(\boldsymbol{w}^{(j)^T} h(\boldsymbol{x}_i)\right)}, \quad k \in \mathcal{K},\ i \in \mathcal{S} \tag{2}$$

where $\boldsymbol{w} = [\boldsymbol{w}^{(1)^T}, \ldots, \boldsymbol{w}^{(K)^T}]$ is the feature-weight vector and $h(\boldsymbol{x}) = [h_1(\boldsymbol{x}), \ldots, h_l(\boldsymbol{x})]$ is a vector of l fixed functions of the input. The usual choices for $h(\boldsymbol{x})$ are kernels $h(\boldsymbol{x}) = [1, K(\boldsymbol{x}, \boldsymbol{x}_1), \ldots, K(\boldsymbol{x}, \boldsymbol{x}_n)]^T$, where $K(\cdot, \cdot)$ is some symmetric kernel function. Here, we use the radial basis function (RBF) $K(x_i, x_j) = \exp(-\|x_i - x_j\|^2/2\sigma^2)$. Assuming the sparsity of regressors $\boldsymbol{w}$, the general prior of $\boldsymbol{w}$ is modeled as Laplacian density [9]

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

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

The MAP estimation of the regressors can be written as

$$\widehat{\boldsymbol{w}} = \arg\max_{w} l(\boldsymbol{w}) + \log p(\boldsymbol{w}) \tag{4}$$

where $l(\boldsymbol{w})$ is the log-likelihood function

$$l(\boldsymbol{w}) = \log \prod_i p(y_i|\boldsymbol{x}_i, \boldsymbol{w}). \tag{5}$$

The problem (4) is concave and difficult to solve. However, the log-likelihood function (5) can be approximated to a quadratic lower bound [7]. Based on this lower bound, problem (4) becomes convex and easy to solve. Therefore, the method LORSAL[1] proposed in [10] is good at inferring the SMLR regressors given in (4) for hyperspectral data problem, and its complexity is $O(L^2(K-1))$.

[1] http://www.lx.it.pt/jun/

As the SMLR only considers the spectral information, the combination of SMLR and MRF is proposed to impose both spectral information and spatial information [24] in Bayesian framework. In [24], the Gibbs distribution which specifies a certain MRF [31] is adopted to model the spatial prior $p(\mathbf{y})$, and this leads to a difficult integer optimization problem. Although this kind of integer optimization problem can be approximately solved by α-expansion algorithm [38], it is time-consuming. In this paper, the proposed HSI classification will be treated as an adaptive total variation (TV) regularized SMLR model which can be derived from a weighted MRF under Bayesian inferring framework. The TV regularization can be efficiently solved by ADMM, and the advantages are the fact that the minimizing problem can be transformed into several subproblems, and each of them has a closed-form solution.

B. Proposed SMLR-SpATV Framework

In this paper, from the Bayesian inferring perspective, the densities $p(y_i|\boldsymbol{x}_i)$ are modeled as SMLRs according to Li's work [24], and the prior $p(\mathbf{y})$ is not directly modeled as a Gibbs distribution on the discrete labels $\mathbf{y}$. Instead, the spatial prior is modeled on the implicit marginal probability of the posterior distribution, as the classification labels $\mathbf{y}$ are totally determined by their corresponding marginal probabilities (the details would be described in the following section). The framework of our method is illustrated in Fig. 1. The main contributions in this paper can be summarized as follows.

1) The spatial prior is modeled as a weighted MRF on the implicit marginal probability instead of the labels $\mathbf{y}$, and the weights are calculated by the gradients of the original HSI.
2) The Bayesian classification model is transformed into a spatially adaptive TV regularization model, in which the marginal probabilities $p(y_i|\boldsymbol{x}_j, j \in N_i)$ (N_i is the set of the nearest neighbors of i) are associated with the spatial prior that neighboring pixels in HSI are likely to belong to the same class. Moreover, the new model can be easily solved by ADMM.
3) In order to impose the spatial information of training samples in the image, during the iteration of the ADMM algorithm, we reset their 1-of-K encodings to spread their class information to their neighbors. It will result in more stable and accurate classification maps.

III. Proposed Model and Algorithm

A. New MPM Classification Model Using Spatially Adaptive TV

MAP marginal (MPM) solution is an alternative to the MAP solution in HSI classification [26]. The MPM estimate of label y_i is given by

$$\widehat{y}_i \equiv \max_k q_i^{(k)} = \max_k p(y_i = k|\boldsymbol{x}_j, j \in N_i), \quad i \in \mathcal{S}. \tag{6}$$

Therefore, our task is to design an algorithm to obtain the marginal solution of MAP. Meanwhile, we can impose the spatial prior on the implicit marginal distribution $\mathbf{q}$ instead of the

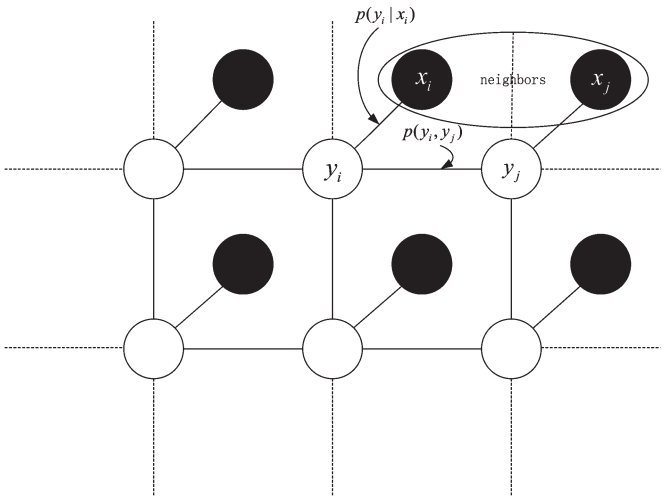


Fig. 2. Graphical example of an MRF.

classification labels $\mathbf{y}$. For convenience, we note the probability matrix gotten by SMLR as $\mathbf{p}_{\mathrm{SMLR}}$ and the implicit marginal matrix as $\mathbf{q}$, where $\mathbf{p}_{\mathrm{SMLR}} \equiv [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_N] \in R^{K\times N}$ with $\boldsymbol{p}_i = [p_i^{(1)}, \ldots, p_i^{(K)}]^T$ and $\mathbf{q} \equiv [\boldsymbol{q}_1, \ldots, \boldsymbol{q}_N] \in R^{K\times N}$ with $\boldsymbol{q}_i = [q_i^{(1)}, \ldots, q_i^{(K)}]^T$.

By using the spectral-only classifier SMLR to obtain the initial probability map $p_{\mathrm{SMLR}} = p(y_i|\boldsymbol{x}_i)$, there will often be many classification errors or outliers in the homogeneous areas due to the lack of spatial contextual constraint. By viewing the initial SMLR error as random additive noise, we can model the spectral-only probability map as

$$\mathbf{p}_{\mathrm{SMLR}} = \mathbf{q} + \mathbf{v} \tag{7}$$

where $\mathbf{p}_{\mathrm{SMLR}}$ is the initial spectral-only probability obtained by SMLR, $\mathbf{q}$ is the unknown marginal probability, and $\mathbf{v}$ is the SMLR outliers. It is well known that inferring $\mathbf{q}$ from $\mathbf{p}_{\mathrm{SMLR}}$ is an ill-posed inverse problem, and MRF-based regularization is usually used to find a stable solution.

In the Bayesian framework, inferring $\mathbf{q}$ from $\mathbf{p}_{\mathrm{SMLR}}$ is defined as maximizing the posterior probability $p(\mathbf{q}|\mathbf{p}_{\mathrm{SMLR}})$. By Bayes formula

$$\begin{aligned}\mathbf{q} &= \arg\max_{\mathbf{q}} p(\mathbf{q}|\mathbf{p}_{\mathrm{SMLR}}) \\ &= \arg\max_{\mathbf{q}} p(\mathbf{p}_{\mathrm{SMLR}}|\mathbf{q})p(\mathbf{q})/p(\mathbf{p}_{\mathrm{SMLR}}).\end{aligned} \tag{8}$$

Once $\mathbf{p}_{\mathrm{SMLR}}$ is given, $p(\mathbf{p}_{\mathrm{SMLR}})$ is a normalization constant. In essence, we are then seeking to maximize the product of the data model $p(\mathbf{p}_{\mathrm{SMLR}}|\mathbf{q})$ and the spatial prior on the real-valued marginal distribution $p(\mathbf{q})$. For the data model, we model the SMLR outliers as additive Gaussian white noise, i.e., $p(\mathbf{p}_{\mathrm{SMLR}}|\mathbf{q}) \propto \exp(-\|\mathbf{p}_{\mathrm{SMLR}} - \mathbf{q}\|_F)$, where $\|\cdot\|_F$ denotes the Frobenius norm. For the prior term, we impose the spatial prior that the neighboring pixels in hyperspectral scene are likely to belong to the same class by $p(\mathbf{q})$. According to the MRF (see Fig. 2) spatial prior in [25], it can be modeled as

$$p(\mathbf{q}) = \frac{1}{Z}\exp\left\{-\mu_s \sum_{|i-j|<\delta} \|\boldsymbol{q}_i - \boldsymbol{q}_j\|_1\right\} \tag{9}$$

where Z is a normalization constant for the density and $|i-j| < \delta$ denotes that pixel i and pixel j are a pair of neighbors. The parameter μ_s tunes the degree of homogeneity of each region in the image. As illustrated in [32], the value of μ_s plays an important role in the number and the size of the regions. Moreover, its value deeply depends on the neighborhood structure [33]. In order to impose the spatial structure information in the original hyperspectral data, we use a spatially adaptive parameter to compromise the two terms. Then, the prior can be written as

$$p(\mathbf{q}) = \frac{1}{Z}\exp\left\{-\mu_s \cdot a(x_i) \sum_{|i-j|<\delta} \|\boldsymbol{q}_i - \boldsymbol{q}_j\|_1\right\} \tag{10}$$

where $a(\boldsymbol{x}_i)$ is the spatially adaptive parameter imposing the edge information, and it will be declared in the following section.

At this point, the solution to (8) corresponding to the MAP is equivalent to the solution that minimizes the following relations:

$$\begin{aligned}\mathbf{q} &= \arg\max_{\mathbf{q}} p(\mathbf{q}|\mathbf{p}_{\mathrm{SMLR}}) \\ &= \arg\min_{\mathbf{q}} \left(-\log p(\mathbf{p}_{\mathrm{SMLR}}|\mathbf{q}) - \log p(\mathbf{q})\right).\end{aligned} \tag{11}$$

Substituting $p(\mathbf{p}_{\mathrm{SMLR}}|\mathbf{q})$ and $p(\mathbf{q})$ into (11), we can derive a regularized estimate of $\mathbf{q}$ which can be generally written as

$$\widehat{\mathbf{q}} = \arg\min_{\mathbf{q}}\left\{\|\mathbf{p}_{\mathrm{SMLR}} - \mathbf{q}\|_F^2 + \mu_s \cdot a(\boldsymbol{x}_i) \sum_{|i-j|<\delta} \|\boldsymbol{q}_i - \boldsymbol{q}_j\|_1\right\}. \tag{12}$$

The first term is the data fidelity term penalizing the inconsistency between $\mathbf{q}$ and the initial SMLR-based spectral-only probability map $\mathbf{p}_{\mathrm{SMLR}}$, and the second one is the regularization term depending on the spatial smoothness prior knowledge on $\mathbf{q}$.

As $\mathbf{q}$ is the probability distribution, it must be nonnegative, and the sum of its column must be 1. Meanwhile, in order to use the spatial information of training samples, we fix their 1-of-K encodings in $\mathbf{q}$. The final proposed model using spatial–spectral information can be formulated by

$$\begin{aligned}&\widehat{\mathbf{q}} = \arg\min_{\mathbf{q}} \frac{1}{2}\|\mathbf{q} - \mathbf{p}_{\mathrm{SMLR}}\|_F^2 + \lambda_{\mathrm{TV}}\mathrm{TV}(\mathbf{q}) \\ &\text{s.t.} \quad \mathbf{q} \geq 0, \quad \mathbf{q}_{\Lambda_l} = \mathbf{y}_{\Lambda_l}, \quad 1^T\boldsymbol{q}_i = 1, \quad i = 1, \ldots, N\end{aligned} \tag{13}$$

where $\|\cdot\|_F$ is the Frobenius norm, $\lambda_{\mathrm{TV}} = \mu_s \cdot a(\mathbf{x}) \geq 0$ are spatially adaptive regularization parameters, Λ_l is the index set of training samples, and $\mathbf{y} \equiv (y_1, y_2, \ldots, y_N) \in \mathcal{K}^N$ is the image of class labels with 1-of-K encodings. If $\boldsymbol{x}_i$ is a training sample of class k, y_i is a K-dimensional vector, with the kth element being 1 while the other elements being zeroes. By doing this, the labels of the training samples will spread their labels from their positions to the whole data until a stable status is achieved, and it effectively improves the classification result.

According to [39], let $\mathbf{H}_h : \mathbb{R}^{K\times N} \to \mathbb{R}^{K\times N}$ denote a linear operator computing the horizontal differences between the components of $\mathbf{q}$ corresponding to neighboring pixels, i.e., $\mathbf{H}_h\mathbf{q} = [\boldsymbol{d}_1, \ldots, \boldsymbol{d}_N]$, where $\boldsymbol{d}_i = \boldsymbol{q}_i - \boldsymbol{q}_{i_h}$ with i and i_h denoting a pixel and its horizontal neighbor. Let $\mathbf{H}_v$ be defined in a similar way for the vertical differences. We have

$$\mathbf{Hq} \equiv \begin{bmatrix}\mathbf{H}_h\mathbf{q}\\ \mathbf{H}_v\mathbf{q}\end{bmatrix}.$$

With these definitions in place, the spatial prior can be rewritten as

$$\mathbf{TV}(\mathbf{q}) \equiv \mathbf{Hq} = \sum_{|i-j|<\delta} \|\boldsymbol{q}_i - \boldsymbol{q}_j\|_1 \tag{14}$$

where $\mathbf{H}$ is a convolution operator, $\boldsymbol{q}_i$ is the ith column of $\mathbf{q}$ corresponding to pixel i, $\|\cdot\|_1$ denotes the l_1 norm, $|i-j|<\delta$ denotes that the ith pixel and the jth pixel are neighbors in spatial domain of the hyperspectral data, and δ controls the size of the neighborhood. A larger value of δ would lead to a smoother classification result. It is worth to emphasize that the neighborhood means not only the local neighborhood but also the nonlocal neighborhood, e.g., if $\boldsymbol{q}_{i_h}$ is the mean value of the nonlocal horizontal neighbors of $\boldsymbol{q}_i$, the $\mathbf{TV}$ in this situation can be understood in nonlocal sense. In this paper, we only focus on the local system with first-order neighborhood ($\delta = 1$). This vector extension of the nonisotropic $\mathbf{TV}$ regularizer on the marginal distribution $\mathbf{q}$ would encourage the pixels in the same neighborhood to belong to the same class. Therefore, this regularization would promote a piecewise smooth classification result.

B. Algorithm

ADMM method [35]–[37], [39] is adopted to solve this TV relevant model (13). By adding the variables V_1, V_2, and V_3, model (13) can be rewritten as

$$\begin{aligned}\arg\min_{\mathbf{q}} \Big\{ &\frac{1}{2}\|\mathbf{q}-\mathbf{p}_{\mathrm{SMLR}}\|_F^2 + \lambda_{\mathrm{TV}}\|V_2\|_{1,1} + l_{R^+}(V_3)\\ &+ l_{\{1\}}(1^T\boldsymbol{q}_i)\Big\}\\ \text{s.t.}\quad & V_1 = \mathbf{q}\\ & V_2 = \mathbf{H}V_1\\ & V_3 = \mathbf{q}\end{aligned} \tag{15}$$

where $\|x\|_{1,1} = \sum_{i=1}^{N}\|x_i\|_1$, x_i is the ith column of x, and $l_s(x)$ is equal to 0, while x belongs to set s or $+\infty$ otherwise, and the ADMM algorithm for the problem (15) is shown in **Algorithm 1**, where

$$\begin{aligned}&\mathcal{L}(\mathbf{q}, V_1, \ldots, V_3, D_1, \ldots, D_3)\\ &= \frac{1}{2}\|\mathbf{q}-\mathbf{p}_{\mathrm{SMLR}}\|_F^2 + \lambda_{\mathrm{TV}}\|V_2\|_{1,1} + l_{R^+}(V_3) + l_{\{1\}}(1^T\boldsymbol{q}_i)\\ &\quad + \frac{\mu}{2}\|\mathbf{q}-V_1-D_1\|_F^2 + \frac{\mu}{2}\|\mathbf{H}V_1-V_2-D_2\|_F^2\\ &\quad + \frac{\mu}{2}\|\mathbf{q}-V_3-D_3\|_F^2\end{aligned} \tag{16}$$

and D_1, D_2, and D_3 are Lagrange multipliers, μ is a nonnegative parameter.

Algorithm 1 Split Augmented Lagrange Multiplier Method for SMLR-SpATV.

1: **Initialization** set $t = 0$, $\mu > 0$, V_1^0, V_2^0, V_3^0, D_1^0, D_2^0, D_3^0.
2: **repeat**
3: $\quad\mathbf{q}^{t+1} \leftarrow \arg\min_{q}\ \mathcal{L}(\mathbf{q}, V_1^{(t)}, \ldots, V_3^{(t)}, D_1^{(t)}, \ldots, D_3^{(t)})$
4: $\quad\mathbf{q}_{\Lambda_l}^{t+1} \leftarrow \mathbf{y}_{\Lambda_l}$
5: **for** $i = 1, \ldots, 3$
6: $\quad V_i^{t+1} \leftarrow \arg\min_{V_i} \mathcal{L}(\mathbf{q}, \ldots, V_i^{(t)}, \ldots, D_1^{(t)}, \ldots, D_3^{(t)})$
7: **end for**
8: **Update multipliers**
9: $\quad D_1^{(t+1)} \leftarrow D_1^{(t)} - \mathbf{q}^{(t+1)} + V_1^{(t+1)}$
10: $\quad D_2^{(t+1)} \leftarrow D_2^{(t)} - \mathbf{H}V_1^{(t+1)} + V_2^{(t+1)}$
11: $\quad D_3^{(t+1)} \leftarrow D_3^{(t)} - \mathbf{q}^{(t+1)} + V_3^{(t+1)}$
12: **Update iteration** $t = t + 1$
13: **until some stopping criterion is satisfied**

Step 3 in **Algorithm 1** is a quadratic programming problem. Its solution is

$$\mathbf{q}^{(k+1)} \leftarrow \frac{1}{1+2\mu}\left(\mathbf{p}_{\mathrm{SMLR}} + \mu\left(V_1^{(k)} + D_1^{(k)} + V_3^{(k)} + D_3^{(k)}\right)\right) \tag{17}$$

and the solutions of variables V_1, V_2, and V_3 in step 6 are given as

$$V_1^{(k+1)} \leftarrow (\mathbf{H}^T\mathbf{H}+\mathbf{I})^{-1}\left(\mathbf{H}^T\left(V_2^{(k)} + D_2^{(k)}\right) + \mathbf{q}^{(k+1)} - D_1^{(k)}\right), \tag{18}$$

$$V_2^{(k+1)} \leftarrow soft\left(D_2^{(k)} - \mathbf{H}V_1^{(k)}, \lambda_{\mathrm{TV}}/\mu\right) \tag{19}$$

$$V_3^{(k+1)} \leftarrow \max\left(\mathbf{q}^{(k+1)} - D_3^{(k)}\right). \tag{20}$$

$\mathbf{H}$ in step 10 acts only on spatial domain and can be applied independently in band-by-band fashion. Furthermore, $\mathbf{H}$ is a convolution and can be computed efficiently using the discrete Fourier transform diagonalization. Step 4 is to fix the class labels of the training samples in the iteration, and the spatial information would be spread from the points of training samples to the whole data by the TV constraint. The model is robust on the value of parameter λ_{TV}; even if we choose a large value for it, the result would not be oversmoothed.

Concerning the computational complexity of the SMLR-spatially adaptive total variation (SpATV) algorithm, the most costly step is the calculus of V_2, which has the order of complexity $O(LN\log N)$, while the others have computational complexity $O(N)$.

C. Spatially Adaptive Regularization Parameters

In this section, we want to impose the edge information to improve the accuracy of HSI classification. The important

problem is how to realize the spatially adaptive aspect in the model, i.e., how to adjust the power of the spatial smoothness in different pixel locations in the original hyperspectral image. To this end, the spatially adaptive mechanism is described as follows.

First, we use the gradient information to characterize the edge structure by the following expression:

$$\nabla \boldsymbol{x}_i = \sqrt{\sum_{j=1}^{L} \left(\left(\nabla^h x_i^j \right)^2 + \left(\nabla^v x_i^j \right)^2 \right)} \tag{21}$$

where $\nabla^h x_i^j$ and $\nabla^v x_i^j$ are the horizontal and vertical first-order gradients of $\boldsymbol{x}_i$ at the jth band. Then, $a(\boldsymbol{x}_i)$ is calculated by

$$a(\boldsymbol{x}_i) = \frac{1}{1 + \nabla \boldsymbol{x}_i}. \tag{22}$$

With the aforementioned definition in (22), it is obvious that, for pixels in smooth regions, the value of $a(\boldsymbol{x}_i)$ will be large; therefore, a strong strength of smoothness will be used for these pixels. Conversely, for the pixels in edge or texture areas, the value of $a(\boldsymbol{x}_i)$ will be small, the weak strength of smoothness will be used for them, and the fine structures in the image will be preserved.

IV. Experimental Results

In this section, we apply the proposed method named SMLR and spatially adaptive TV classifier (SMLR-SpATV) to three hyperspectral airborne images described as follows.

1) The Indian Pines image was collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over the Indian Pines region in June 1992. The AVIRIS sensor generates 220 bands across the spectral range from 0.2 to 2.4 μm. In the experiments, the number of bands is reduced to 200 by removing 20 water absorption bands [3]. This scene is available online[2] and has a spectral resolution of 10 nm and a spatial resolution of 20 m by pixel, and the spatial dimension is 145 $\times$ 145. The ground truth contains 16 land cover classes and a total of 10 366 labeled pixels. The number of pixels in the smallest class is 20, while the number of pixels in the largest class is 2468, as seen in Table I.
2) The University of Pavia image was acquired in 2001 by the Reflective Optics System Imaging Spectrometer (ROSIS), flown over the city of Pavia, Italy. The sensor generates 115 spectral bands ranging from 0.43 to 0.86 μm and has a spatial resolution of 1.3 m per pixel. The image scene, with a size of 610 $\times$ 340 pixels, is centered at the University of Pavia and has 103 bands after removing 12 noisiest bands. There are nine ground truth classes, 3921 (about 9%) of all labeled data are used in training, and the rest is used for testing, as seen in Table II.
3) The center of Pavia image was the other urban image collected by the ROSIS sensor over the center of the Pavia city in 2001. This image consists of 1096 $\times$ 492 pixels, each having 102 spectral bands after 13 noisy bands are removed. The nine ground truth classes and the number of training and test samples for each class are shown in Table III. Here, 5536 labeled samples (about 5% of all of the labeled samples) are used for training purposes, and the rest is for testing.

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

TABLE I
Sixteen Ground Truth Classes in the AVIRIS Indian Pines

Class		samples	
No.	Name	Train	Test
1	Alfalfa	6	48
2	Corn-no till	144	1290
3	Corn-min till	84	750
4	Corn	24	210
5	Grass/pasture	50	447
6	Grass/tree	75	672
7	Grass/pasture-mowed	3	23
8	Hay-windrowed	49	440
9	Oats	2	18
10	Soybeans-no till	97	871
11	Soybeans-min till	247	2221
12	Soybeans-clean till	62	552
13	Wheat	22	190
14	Woods	130	1164
15	Bldg-grass-tree-drives	38	342
16	Stone-steel towers	10	85
Total		1043	9323

TABLE II
Nine Ground Truth Classes in the University of Pavia Image

Class		samples	
No.	Name	Train	Test
1	Asphalt	548	6304
2	Meadows	540	18146
3	Gravel	392	1815
4	Trees	524	2912
5	Metal sheets	265	1113
6	Bare soil	532	4572
7	Bitumen	375	981
8	Bricks	514	3364
9	Shadows	231	795
Total		3921	40002

TABLE III
Nine Ground Truth Classes in the Center of Pavia Image

Class		samples	
No.	Name	Train	Test
1	Water	745	64533
2	Trees	785	5722
3	Meadow	797	2094
4	Bricks	485	1667
5	Soil	820	5729
6	Asphalt	678	6847
7	Bitumen	808	6479
8	Tile	223	2899
9	Shadows	195	1970
Total		5536	97940

The goal of the experiments is to assess the performance of the proposed algorithm in association with the smoothness parameter μ_s, the training samples l, and the RBF kernel parameter σ, whereas the main goal of using the three real data sets is to compare the proposed approach with other state-of-the-art analysis classifiers, such as the l_1 regularized sparse regression method [19], (kernelized) orthogonal matching pursuit (denoted as KOMP and OMP, respectively) [19], (kernelized) simultaneous orthogonal matching pursuit (denoted as KSOMP and SOMP, respectively) [19], SOMP with CK (SOMP-CK)

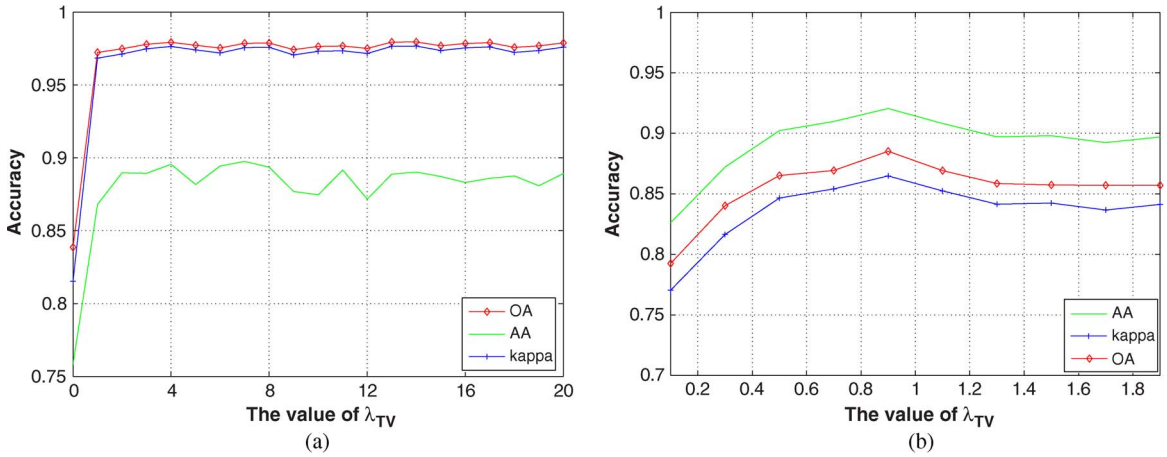


Fig. 3. Impact of the smoothness parameter λ_{TV} in (a) the Indian Pines image and (b) the University of Pavia image.

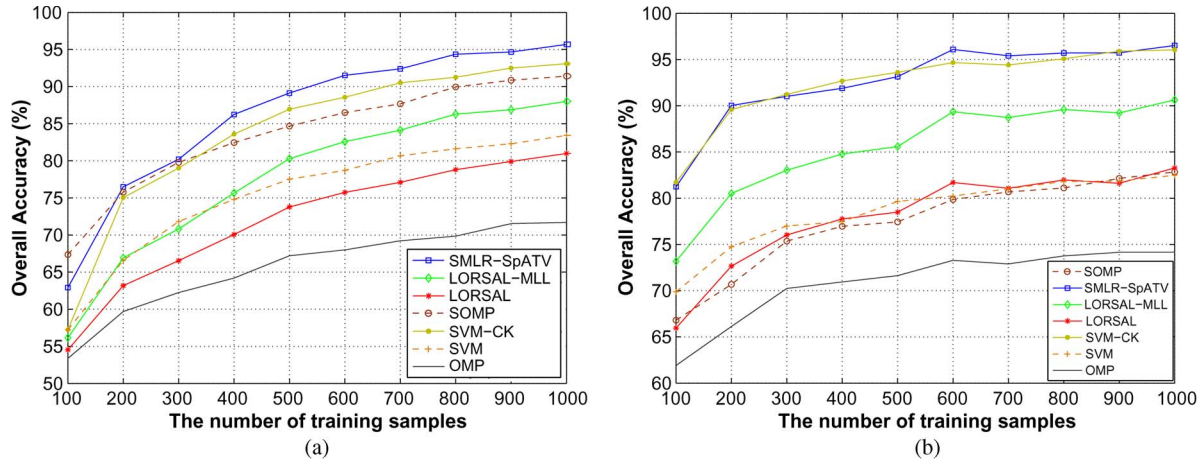


Fig. 4. Impact of the number of training samples (a) the Indian Pines image and (b) the University of Pavia image.

[20], LORSAL [11], LORSAL with multilevel logistic spatial prior (denoted as LORSAL-MLL) [11], LORSAL with CK (denoted as LORSAL-CK) [20], and SVMs [3], [6], which are well-established techniques in the machine learning community. For SVMs, we use a CK (denoted by SVM-CK) [16] that combines the spectral information and spatial information via a weighted kernel summation. Moreover, SVM-CK has proven to outperform the spectral-only SVM in HSI classification. All parameters of these methods are set according to the reference papers. For the OMP, SOMP, KOMP, KOMP-CK, SVM, and SVM-CK algorithms, it is needed to normalize the data set before doing the experiments.[3]

A. Measurements and Parameter Setting

Before doing the experiments, we first describe the parameter setting and the measurements of classifications. In our proposed algorithm, there are three important parameters, namely, μ_s, $a(\boldsymbol{x}_i)$, and the RBF kernel parameter σ. In the experiments, σ is selected according to the cross validation. The parameter $a(\boldsymbol{x}_i)$ is precalculated according to the hyperspectral data itself by formula (22), and before calculating the parameter $a(\boldsymbol{x}_i)$, all data sets are normalized for calculating the gradients of the original hyperspectral data, so its value belongs to the interval [0,1]. Moreover, the smoothness parameter μ_s is important to the result; we will analyze its impact in the following experiments. If no special emphasis, σ is set to 0.85 for Indian Pines, 0.35 for the University of Pavia, and 0.35 for Pavia center, and μ_s is set to 2 for all three real data sets. For the LORSAL algorithm, the regularization parameter λ in (3) controls the sparseness of the regressors. According to [24], λ is set to 10^{-2} for all of the following experiments.

The classification results are measured by the overall accuracy (OA), average accuracy (AA), and kappa statistic (k). The OA is computed by the ratio between the correctly classified test samples and the total number of test samples, and the AA is the mean of each class accuracies. The k statistic is computed by weighting the measured accuracies. It incorporates both of the diagonal and off-diagonal entries of the confusion matrix and is a robust measure of the degree of agreement. All of these measurements reported in the following experiments are achieved by the adopted classifiers after ten Monte Carlo runs.

B. Impact of the Parameters μ_s, l, and σ

In the first experiment, we conduct an evaluation of the impact of the smoothness parameter μ_s, the number of the

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

TABLE IV
CLASSIFICATION ACCURACY (%) FOR THE UNIVERSITY OF PAVIA IMAGE USING TRAINING AND TESTING SAMPLES AS SHOWING IN TABLE II

Class	SVM	SVM-CK	OMP	KOMP	SOMP	KOMP-CK	l_1	LORSAL	LORSAL-MLL	LORSAL-CK	KSSP	SMLR-SpATV
1	84.30	79.85	59.73	76.09	59.33	82.23	80.65	80.88	88.48	77.17	89.56	**94.57**
2	67.01	**84.68**	66.16	69.61	78.15	72.47	64.74	72.63	76.22	81.61	79.98	82.56
3	68.43	81.87	66.51	72.12	83.53	82.26	73.22	70.75	73.56	82.42	**85.45**	81.13
4	97.80	96.36	97.10	98.11	96.91	98.56	98.35	97.52	**98.76**	95.46	98.66	95.01
5	99.37	99.37	99.85	99.73	99.46	99.82	99.91	99.26	99.70	99.03	99.91	**100.0**
6	92.45	93.55	69.38	87.66	77.41	93.92	92.54	94.51	97.47	96.94	95.76	**100.0**
7	89.91	90.21	88.42	88.07	98.57	92.46	86.95	90.83	94.74	93.83	97.96	**99.17**
8	92.42	92.81	79.09	89.51	89.09	78.78	81.54	90.44	96.66	94.65	96.43	**98.45**
9	97.23	95.35	89.97	93.96	91.95	96.98	98.99	96.20	**99.37**	97.47	98.49	95.45
OA	79.15	87.18	71.17	78.33	79.00	81.07	76.87	81.63	85.69	86.16	87.65	**90.01**
AA	87.66	90.47	79.58	86.10	86.04	88.61	86.32	88.11	91.66	90.95	93.58	**94.04**
k	0.737	0.833	0.639	0.725	0.728	0.758	0.709	0.769	0.819	0.824	0.840	**0.872**

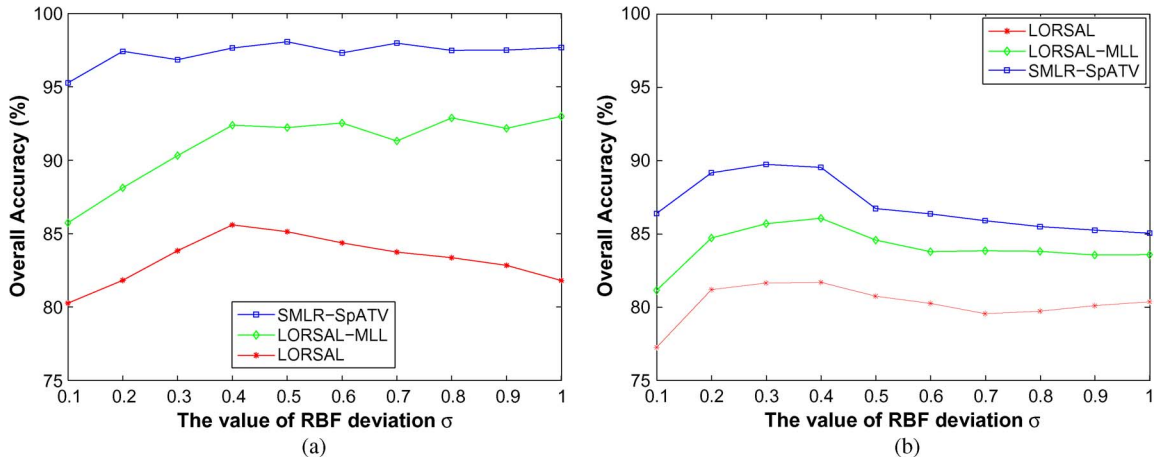


Fig. 5. Impact of the RBF kernel deviation σ (a) the Indian Pines image and (b) the University of Pavia image.

training samples l, and the RBF kernel deviation σ using the Indian Pines and University of Pavia.

Fig. 3(a) and (b) plots the OA, AA, and kappa statistic results as a function of variable μ_s with the training samples of 1043 and 3921 (about 10% and 9% of the available labels, respectively) in the Indian Pines and the University of Pavia, respectively. In each Monte Carlo run, we select the training samples randomly from each class of the available samples of the two data sets. From Fig. 3(a) and (b), we can conclude that the classification performance indeed depends on the value of μ_s. However, in a wide interval value of μ_s, the proposed SMLR-SpATV algorithm leads to good classification results. This indicates that the approach is not very sensitive to the value of μ_s, e.g., even if we choose a value of 10, we can get a good performance, about 97% in OA for the Indian Pines data.

Fig. 4(a) and (b) plots the OA results as a function of the number of training samples l with a suboptimal $\mu_s = 2$ and $\sigma = 0.4$ in the two data sets, respectively. For comparison, we also show the OA of LORSAL and LORSAL-MLL method in this experiment. From Fig. 4(a) and (b), it is obvious that the results obtained by our method are robust even with very limited training samples. For example, even 400 training samples in the Indian Pines image lead to about 90% classification accuracy; this is higher than LORSAL-MLL and LORSAL. As we expect, the number of training samples increases, and the OA increases highly. However, for the University of Pavia data set, it is worth to emphasize that the training samples here are randomly selected from each class in the test samples

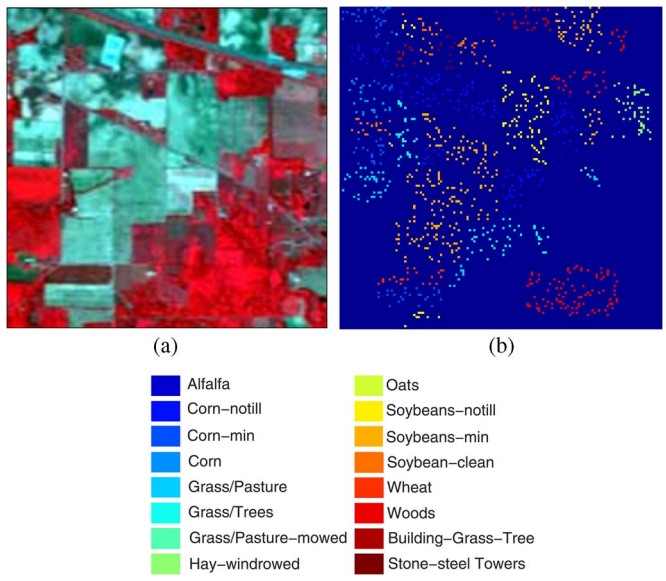


Fig. 6. AVIRIS Indian Pines dataset. (a) False color map. (b) Training data.

shown in Table II. This random selection (RS) will increase the classification accuracy as our fixing the labels of the training samples in our algorithm. Therefore, the OA in Fig. 4(b) is higher than that shown in Table IV.

Fig. 5(a) and (b) plots the OA results as a function of the RBF kernel deviation parameter σ in the Indian Pines and the University of Pavia, respectively. From Fig. 5(a) and (b), we can see that our method gets the best results among the three

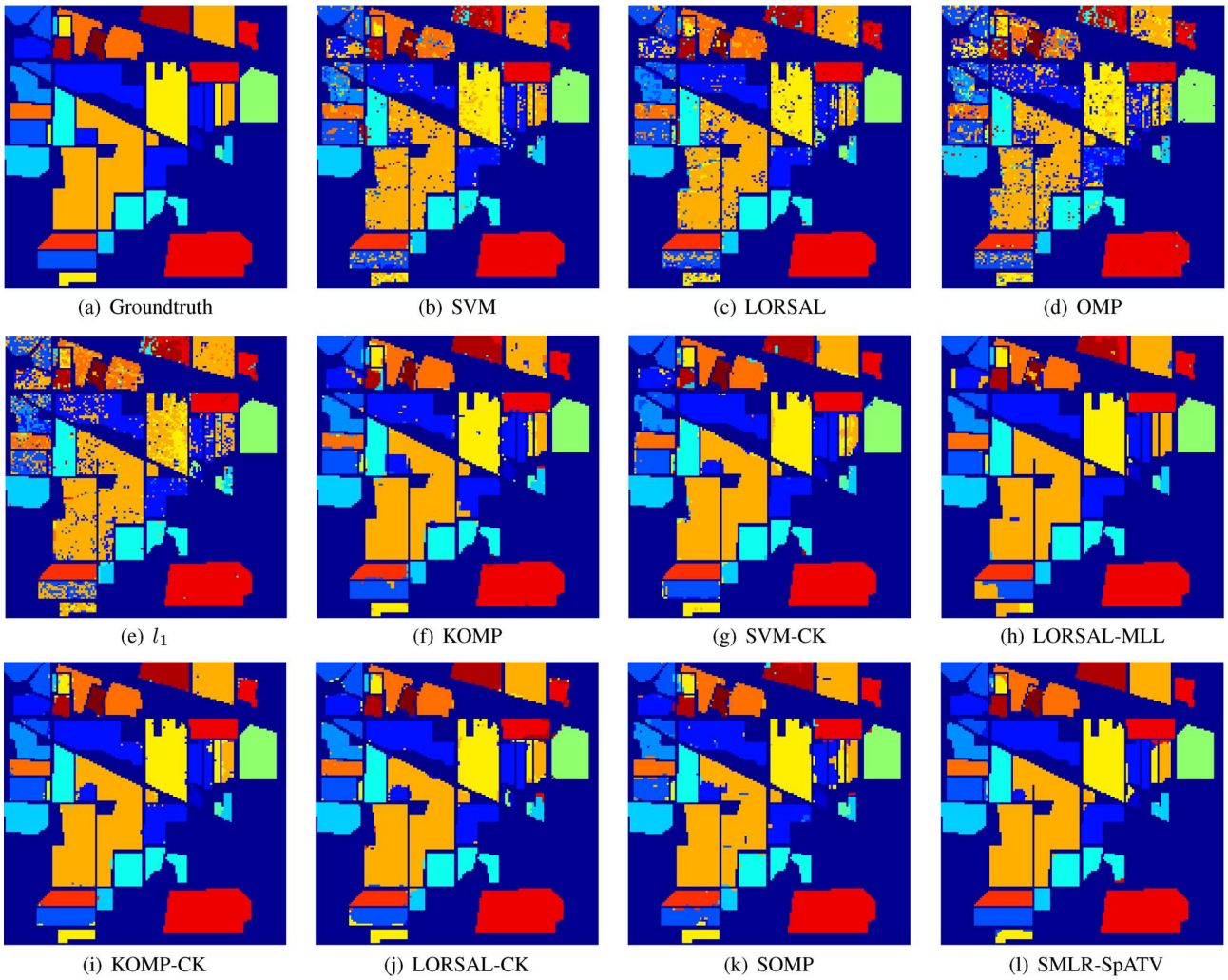


Fig. 7. Indian Pines image. (a) ground truth and the overall accuracy of the methods (b) SVM (OA = 84.52%); (c) LORSAL (OA = 83.94%); (d) OMP (OA = 75.99%); (e) l_1 (OA = 77.99%); (f) KOMP (OA = 85.12%); (g) SVM-CK (OA = 94.86%); (h) LORSAL-MLL (OA = 92.70%); (i) KOMP-CK (OA = 97.42%); (j) LORSAL-CK (OA = 95.81%); (k) SOMP (OA = 95.28%); (l) SMLR-SpATV (OA = 97.85%) with about 10% training samples.

compared classifiers, and even if LORSAL is a little sensitive to the parameter σ, our method is robust on it.

C. AVIRIS Indian Pines Data Set

In this experiment, we show the classification results (OA, AA, kappa statistic coefficient (k), and individual class accuracy) of the proposed method using the real hyperspectral data set Indian Pines. For comparison, we adopt the aforementioned state-of-the-art supervised classifiers.

In order to evaluate the performance of our algorithm with limited training samples, a total size of $l = 1043$ (about 10% of the available samples) training samples was used for training purposes, and the training samples were randomly selected in each class according to Table I, while the rest was used for testing. For illustration, Fig. 6(b) shows the training samples. Fig. 7(a)–(l) shows the ground truth and classification results obtained by the different tested methods for the Indian Pines data. For each method, we randomly choose one of the maps and the AAs obtained by ten Monte Carlo runs. Moreover, Table V gives all comparable results of different classifiers. From Table V, it is obvious that the classifiers with spatial information (MLL prior, CK, or simultaneous sparse representation methods) have shown a clear advantage over the pixel-only counterparts. As in Table V, SVM gives the best pixelwise classification result, while our method SMLR-SpATV gives the best spectral–spatial classification result, which is about 13% higher than the pixelwise result in OA. This is not only due to imposing the spatial information but also to the benefit from fixing the labels of the training samples.

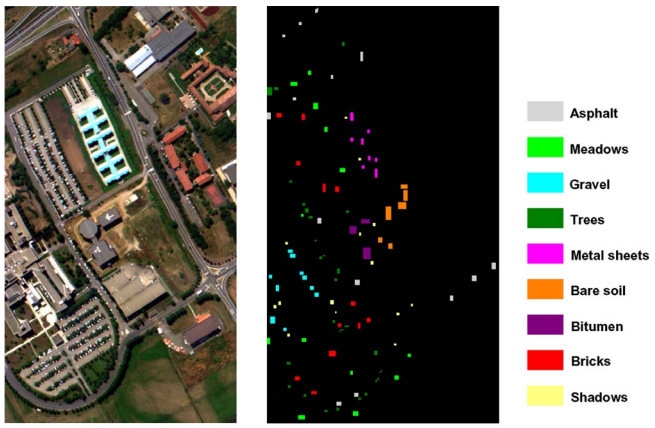
D. ROSIS University of Pavia Data Set

In this experiment, we evaluate our method using the ROSIS University of Pavia data set while comparing with the other state-of-the-art methods mentioned previously. In addition to these aforementioned classifiers, we add a new method proposed recently in [20], named kernelized simultaneous subspace pursuit, for comparison, which obtains the best classification result in the University of Pavia data set in [20].

Fig. 8(b) shows the training samples of the University of Pavia data set. Fig. 9(a)–(l) illustrates the reference map and classification results of the classifiers listed in Table IV. From them, we can conclude that our approach SMLR-SpATV

TABLE V
CLASSIFICATION ACCURACY (IN PERCENT) FOR THE INDIAN PINES IMAGE ON THE TEST SET

Class	SVM	SVM-CK	OMP	KOMP	SOMP	KOMP-CK	l_1	LORSAL	LORSAL-MLL	LORSAL-CK	SMLR-SpATV
1	81.25	**95.83**	63.75	75.00	85.42	94.79	39.58	64.17	70.4167	82.71	91.25
2	86.28	96.67	63.46	81.50	94.88	96.75	78.53	82.03	92.96	96.23	**96.81**
3	72.80	90.93	61.08	74.63	94.93	95.60	51.87	70.97	86.65	95.45	**98.79**
4	58.10	85.71	47.43	69.57	91.43	96.48	28.57	64.81	79.38	94.24	**97.90**
5	92.39	93.74	90.85	91.41	89.49	**97.11**	80.76	91.07	94.25	93.71	95.26
6	96.88	97.32	95.06	96.09	98.51	**99.54**	99.40	96.46	98.66	98.72	98.65
7	43.48	69.57	83.48	76.96	91.30	**94.35**	17.39	38.70	50.00	84.35	60.87
8	98.86	98.41	97.55	98.89	99.55	99.73	99.32	99.30	99.39	99.18	**100**
9	50.00	55.56	53.89	73.89	0	**95.56**	16.67	28.89	50.00	74.44	12.78
10	71.53	93.80	72.35	77.21	89.44	95.06	63.95	76.61	90.20	91.73	**95.88**
11	84.38	94.37	77.61	85.98	97.34	97.74	86.04	83.03	93.89	96.11	**98.98**
12	85.51	93.66	55.78	81.32	88.22	96.99	57.79	81.76	94.37	94.98	**99.11**
13	100.0	99.47	97.42	99.11	**100.0**	99.58	100.0	99.63	99.58	99.68	99.74
14	93.30	99.14	93.99	96.07	99.14	99.05	97.94	96.41	97.66	97.52	**99.51**
15	64.91	87.43	46.17	60.00	**99.12**	96.99	35.96	66.61	79.30	95.56	98.45
16	88.24	**100.0**	89.41	94.00	96.47	94.47	90.59	70.24	73.06	90.59	86.47
OA	84.52	94.86	75.99	85.12	95.28	97.42	77.99	83.94	92.70	95.81	**97.85**
AA	79.24	90.73	74.33	83.23	88.45	96.86	65.27	75.67	84.36	**92.83**	89.40
k	0.823	0.941	0.726	0.830	0.946	0.971	0.746	0.816	0.917	0.952	**0.976**



(a) (b)

Fig. 8. ROSIS University of Pavia dataset. (a) False color map. (b) Training data.

achieves the highest accuracy among all of the other classifiers. It leads to a high classification accuracy, even though the training samples are out of the test samples, which means that the positions of the training samples will be weak when we fix their true labels in our algorithm.

E. ROSIS Center of Pavia Data Set

In this experiment, we evaluate our method using the ROSIS center of Pavia data set while comparing with the other state-of-the-art methods mentioned previously. As in the aforementioned section, we also add a new method proposed in [20] and [40] named KLR-CK, (but here we denote it as LORSAL-CK) which obtains the best classification result in the center of Pavia data set in [20].

Fig. 10(b) shows the training samples of the center of Pavia data set. It is worth to emphasize that these training samples are out of the testing samples. Fig. 11(a)–(l) illustrates the reference map and classification results of the classifiers listed in Table VI. For this data set, all classifiers achieve good classification results, but our method still gets the highest result, which is 0.31% higher than the second best result gotten by LORSAL-CK.

F. Execution Time Comparison

In this section, we compare the execution times of those aforementioned methods. The algorithms are implemented using MATLAB 2010b on a workstation PC equipped with an Intel Xeon CPU (at 3.2 GHZ) and 10 GB of RAM memory. All compared algorithms are programmed in MATLAB. For SVMs, we use the one-versus-one method for multiclass problem.

Table VII reports the average execution times after applying the considered techniques on the Indian Pines with 1043 training samples and the University of Pavia with 2000 of 3921 training samples in 10 Monte Carlo runs, respectively. From Table VII, it is obvious that the classifiers with spatial information cost more time than their pixel-only counterparts. For Indian Pines, LORSAL gets the least execution time among the pixel-only classifiers, while LORSAL-CK gets the least execution time among the spectral–spatial classifiers, and our method gets the third least execution time among the spectral–spatial classifiers. For University of Pavia, LORSAL gets the least execution time among the pixel-only classifiers, while LORSAL-CK again gets the least execution time among the spectral–spatial classifiers, but our method gets the second least execution time among the spectral–spatial classifiers.

G. AL Comparison

In this experiment, we evaluate our SMLR-SpATV method using active learning (AL) [26], [29], [34] to cut the expensive cost spent on collecting the training samples. We consider three different sampling methods for the AL step: 1) an RS, where the new samples are randomly selected from the candidate set; 2) a mutual information (MI) based criterion, which aims at finding the samples maximizing the MI between the MLR regressors and the class labels; 3) a breaking ties (BT) algorithm, which aims at finding the samples minimizing the distance between the first two most probable classes; and 4) a modified break ties (BT) algorithm, which aims at finding samples maximizing the probability of the large class for each individual class.

Our method SMLR-SpATV is based on the Bayesian framework by estimating the MAP using a new spatially adaptive

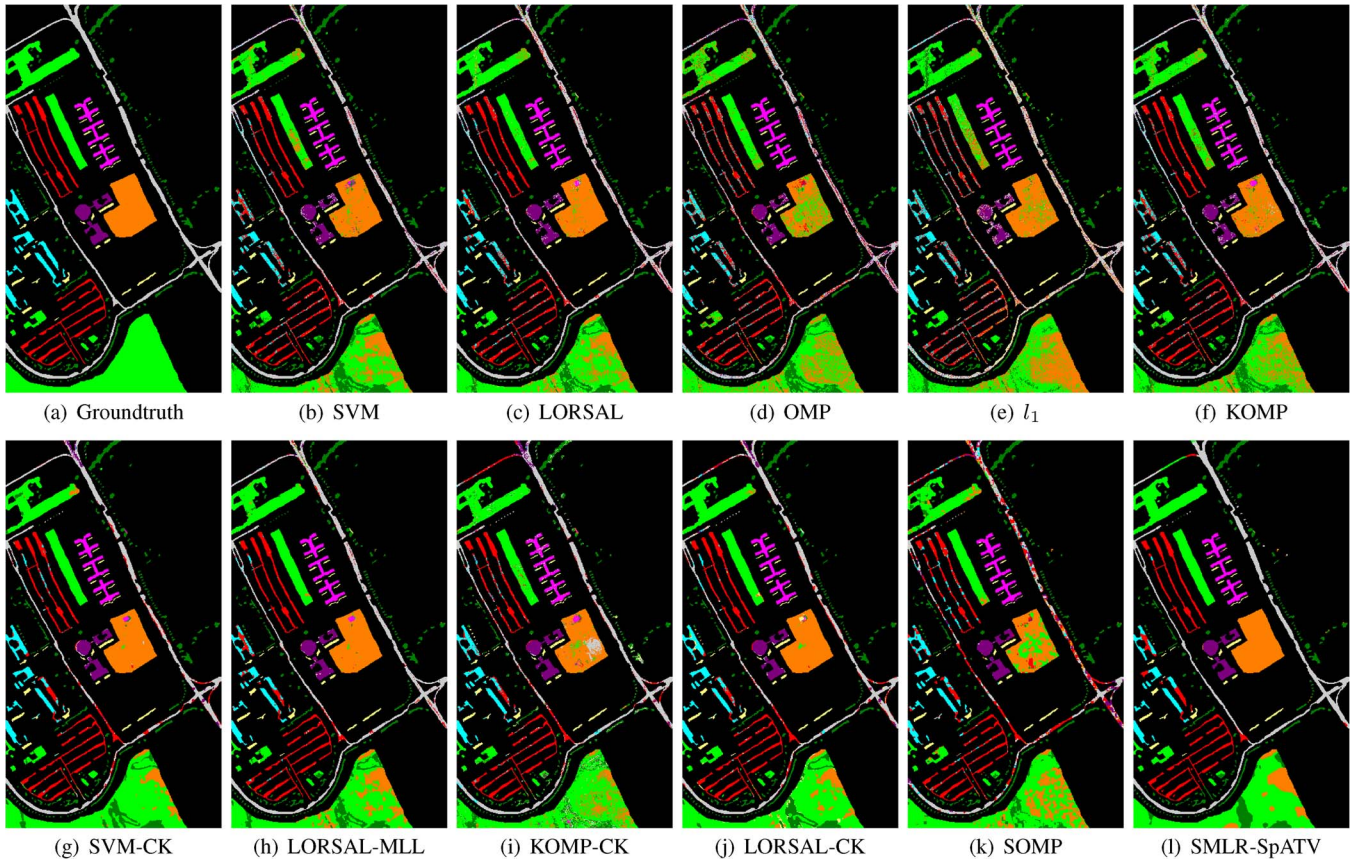


Fig. 9. University of Pavia image. (a) ground truth and the overall accuracy of the methods (b) SVM (OA $= 79.15\%$); (c) LORSAL (OA $= 81.63\%$); (d) OMP (OA $= 71.17\%$); (e) l_1 (OA $= 76.87\%$); (f) KOMP (OA $= 78.33\%$); (g) SVM-CK (OA $= 87.18\%$); (h) LORSAL-MLL (OA $= 85.69\%$); (i) KOMP-CK (OA $= 81.07\%$); (j) LORSAL-CK (OA $= 86.16\%$); (k) SOMP (OA $= 79.00\%$); (l) SMLR-SpATV (OA $= 90.01\%$) with about 9% training samples.

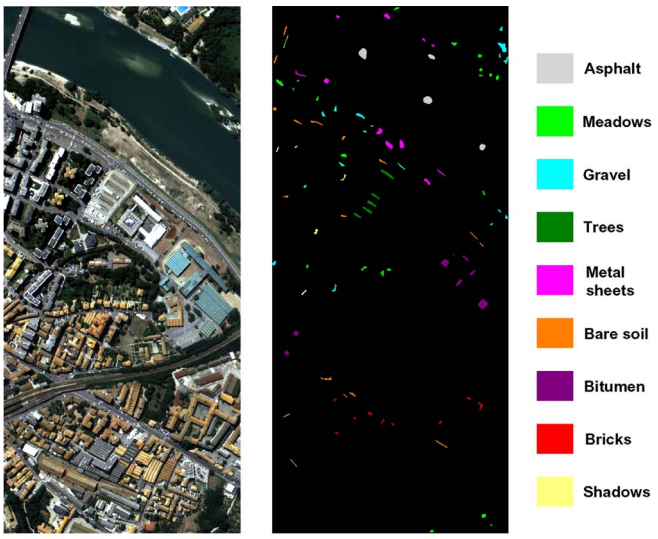


Fig. 10. ROSIS center of Pavia dataset. (a) False color map. (b) Training data.

TV model. As we consider the $p(y_i|\boldsymbol{x}_j, j \in N_i)$ with spatial information using Markov property, the AL will also include the spatial information when applied to the estimated posterior probability.

Fig. 12 shows the results of our proposed SMLR-SpATV-AL, LORSAL-AL, and LORSAL-AL-MLL in the Indian Pines image with an initial training set of 80 samples and ten new samples added at each AL iteration until the training number is up to 610. Fig. 13 shows the results of the AL when the training number is 610. From this two figures, we can conclude that our method makes the best performance, i.e., AL with spatial information has a better performance than pixel-only AL. Moreover, AL is really an effective strategy to reduce the cost of acquiring large labeled training sets in HC. By AL, with only 610 samples of Indian Pines, our method leads to a 99.22% classification accuracy, and it is higher than the result reported in Fig. 4(a).

V. Conclusion

In this paper, we have proposed a new approach for hyperspectral image classification using spectral–spatial information under Bayesian inferring framework. First, the SMLR method is used to learn a spectral-only classifier on the original hyperspectral data. Second, the spatial prior is modeled on the hidden marginal probability of the posterior distribution using an MRF with an adaptive-TV-induced potential function, which encourages neighboring pixels to belong to the same class. The MAP model is finally transformed into a regularized variation model, in which the marginal probability of the posterior distribution is an approximation to the probability learned by SMLR, while the regularization is a spatially adaptive TV. In order to impose the structure of the original data, we use the gradient information to calculate a spatially adaptive parameter to balance the power of smoothness. Furthermore, the positions of training samples are used to spread the labeling information

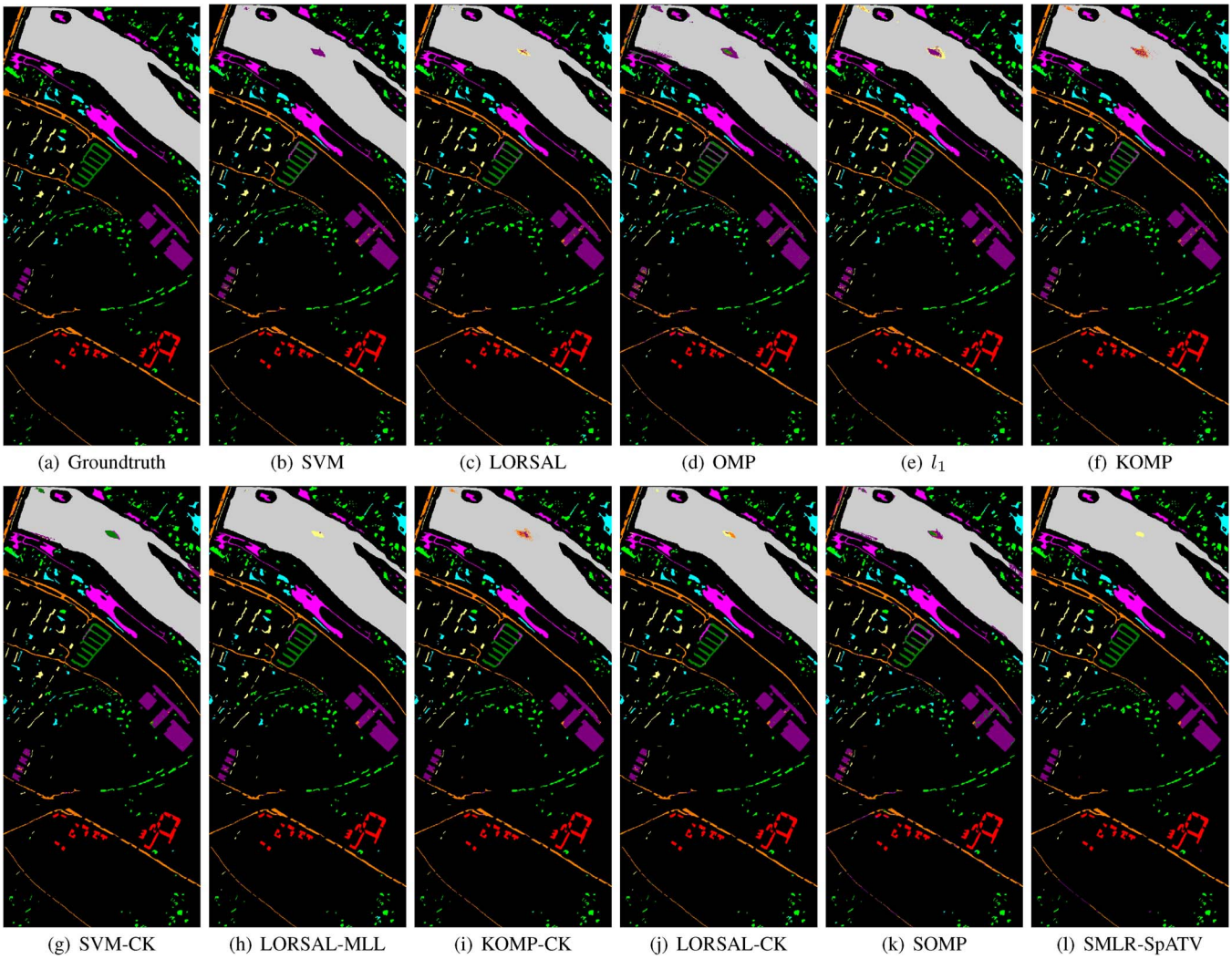


Fig. 11. Center of Pavia image. (a) Ground truth and the OA of the methods. (b) SVM (OA = 97.70%). (c) LORSAL (OA = 91.28%). (d) OMP (OA = 93.63%). (e) l_1(OA = 97.13%). (f) KOMP (OA = 97.37%). (g) SVM-CK (OA = 98.47%). (h) LORSAL-MLL (OA = 98.73%). (i) KOMP-CK (OA = 98.25%). (j) LORSAL-CK (OA = 98.92%). (k) SOMP (OA = 97.66%). (l) SMLR-SpATV (OA = 99.23%) with about 5% training samples. (a) Ground truth. (b) SVM. (c) LORSAL. (d) OMP. (e) l_1. (f) KOMP. (g) SVM-CK. (h) LORSAL-MLL. (i) KOMP-CK. (j) LORSAL-CK. (k) SOMP. (l) SMLR-SpATV.

TABLE VI
CLASSIFICATION ACCURACY (%) FOR CENTER OF PAVIA IMAGE ON THE TEST SET

Class	SVM	SVM-CK	OMP	KOMP	SOMP	KOMP-CK	l_1	LORSAL	LORSAL-MLL	LORSAL-CK	SMLR-SpATV
1	99.23	99.34	96.89	98.81	99.32	99.11	99.35	99.41	99.48	100.0	**99.69**
2	92.64	94.39	81.86	92.96	92.38	96.28	83.52	94.04	95.63	95.39	**98.36**
3	95.51	95.65	95.56	95.46	95.46	**97.61**	97.09	95.18	96.80	95.89	96.80
4	81.64	84.70	74.51	76.90	85.66	83.50	73.67	82.36	89.62	89.80	**93.88**
5	96.04	99.04	91.81	96.16	96.37	98.05	95.92	96.75	97.85	98.59	**99.72**
6	94.99	96.39	78.93	94.27	92.83	95.27	97.01	94.51	97.49	96.67	**97.01**
7	94.61	98.41	91.48	95.40	94.68	97.73	92.65	95.45	98.12	97.31	**99.10**
8	99.72	99.97	97.21	99.52	99.69	99.93	99.45	99.62	99.83	98.41	**100.0**
9	99.95	98.73	93.10	99.95	98.68	99.09	99.09	99.95	**100.0**	99.49	99.34
OA	97.70	98.47	93.63	97.37	97.66	98.25	97.13	97.98	98.73	98.92	**99.23**
AA	94.93	96.43	89.04	94.38	95.01	96.29	93.08	96.32	97.20	96.84	**98.21**
k	0.958	0.972	0.886	0.952	0.958	0.968	0.948	0.953	0.977	0.980	**0.986**

TABLE VII
AVERAGE EXECUTION TIMES (IN SECONDS) AFTER APPLYING THE CONSIDERED TECHNIQUES ON INDIAN PINES WITH 1043 TRAINING SAMPLES AND THE UNIVERSITY OF PAVIA WITH 2000 OF 3921 TRAINING SAMPLES IN TEN MONTE CARLO RUNS, RESPECTIVELY

Data set	SVM	SVM-CK	OMP	KOMP	SOMP	KOMP-CK	l_1	LORSAL	LORSAL-CK	SMLR-SpATV
Indian Pines Time[s]	62.17	63.35	6.44	91.51	96.78	93.78	828.32	6.41	7.26	68.14
PaviaU Time[s]	823.54	833.04	44.06	472.66	1409.8	436.78	1699.6	40.18	58.13	362.25

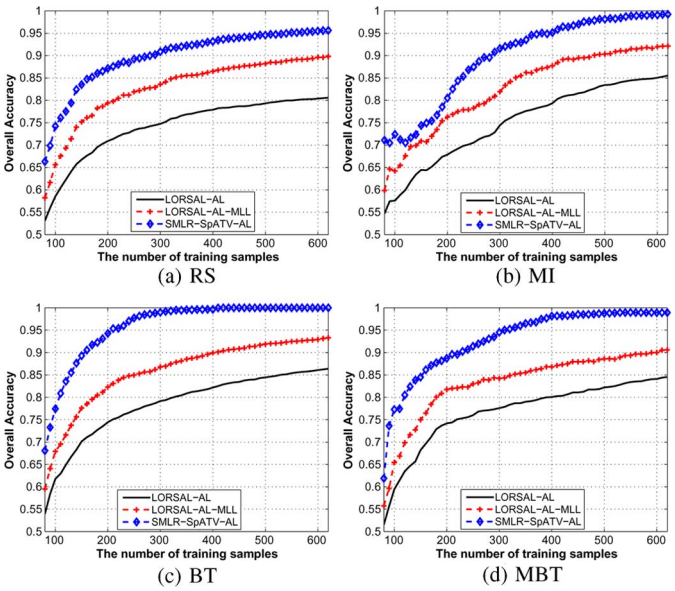


Fig. 12. OA classification results (as a function of the number of labeled training samples) achieved by the LORSAL-AL, LORSAL-AL-MLL, and SMLR-SpATV-AL for the AVIRIS Indian Pines scene. Here, an initial training set of 80 samples and ten new samples are added at each AL iteration until the total number is 620. (a) RS. (b) MI. (c) BT. (d) MBT.

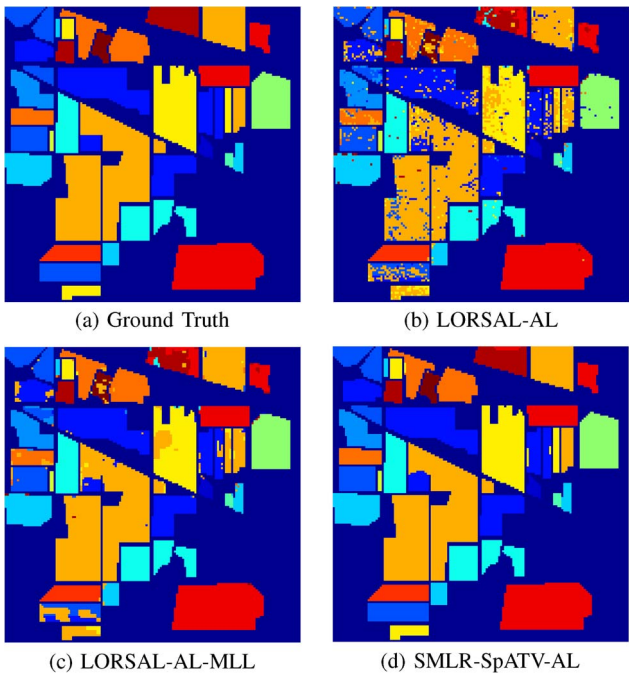


Fig. 13. Indian Pines image. (a) Ground truth and the OA of methods of (b) LORSAL-AL (OA $= 85.55\%$), (c) LORSAL-AL-MLL (OA $= 92.15\%$), and (d) SMLR-SpATV-AL (OA $= 99.22\%$) with about 0.77% (80) initial training samples and adding 10 samples each time by spatial-based AL until 620. (a) Ground truth. (b) LORSAL-AL. (c) LORSAL-AL-MLL. (d) SMLR-SpATV-AL.

from themselves to the whole data to improve the classification result. The ADMM method provides an efficient and fast tool to solve the proposed model with complexity of $O(LNlogN)$. The experimental results of real hyperspectral data sets validate that our proposed method outperforms the other state-of-the-art methods. By the spatial smooth distribution, AL techniques play better than the spectral-only distribution.

The future work will focus on learning the dictionary of each class in the spectral space or its feature space to improve the classification accuracy. Moreover, we will also develop computationally efficient implementations of the proposed approach by resorting to graphical processing units.

ACKNOWLEDGMENT

The authors would like to thank Prof. P. Gamba for providing the ROSIS data over Pavia, Italy, along with the training and testing samples and the 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] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1, pp. 55–63, Jan. 1968.
[2] D. Landgrebe, "Hyperspectral image data analysis as a high dimensional signal processing problem," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.
[3] J. A. Gualtieri and R. F. Cromp, "Support vector machines for hyperspectral remote sensing classification," in *Proc. SPIE*, 1999, vol. 3584, pp. 221–232.
[4] 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.
[5] 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.
[6] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, no. S1, pp. S110–S122, Sep. 2009.
[7] D. Bohning, "Multinomial logistic regression algorithm," *Ann. Inst. Stat. Math.*, vol. 44, no. 1, pp. 197–200, Mar. 1992.
[8] P. Zhong and R. Wang, "Modeling and classifying hyperspectral imagery by CRFs with sparse higher order potentials," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 688–705, Feb. 2011.
[9] B. Krishnapuram, L. Carin, M. A. Figueiredo, and A. J. 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.
[10] J. M. Bioucas-Dias and M. Figueiredo, "Logistic Regression via Variable Splitting and Augmented Lagrangian Tools," Inst. Superior Técnico, TU Lisbon, Lisbon, Portugal, 2009.
[11] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Exploiting spatial information in semi-supervised hyperspectral image segmentation," in *Proc. WHISPERS*, Reykjavik, Iceland, Jun. 2010, pp. 1–4.
[12] J. S. Borges, A. R. Marçal, and J. M. Bioucas-Dias, "Evaluation of Bayesian hyperspectral image segmentation with a discriminative class learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Barcelona, Spain, Jul. 2007, pp. 3810–3813.
[13] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image classification using soft sparse multinomial logistic regression," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 2, pp. 318–322, Mar. 2013.
[14] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using watershed transformation," *Pattern Recognit.*, vol. 43, no. 7, pp. 2367–2379, Jul. 2010.
[15] Y. Tarabalka, J. A. Benediktsson, J. Chanussotand, and J. C. Tilton, "Multiple spectral–spatial classification approach for hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4122–4132, Nov. 2010.
[16] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Marí, J. Vila-Francés, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.
[17] G. Camps-Valls, T. Bandos Marsheva, and D. Zhou, "Semi-supervised graph-based hyperspectral image classification," *IEEE Trans. Geosci. Remote. Sens.*, vol. 45, no. 10, pp. 3044–3054, Oct. 2007.
[18] G. Camps-Valls, N. Shervashidze, and K. M. Borgwardt, "Spatio-spectral remote sensing image classification with graph kernels," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 741–745, Oct. 2010.

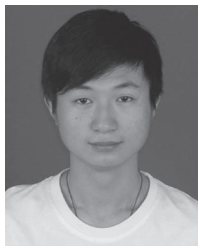
[19] Y. Chen and N. M. Nasrabadi, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote. Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.
[20] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification via kernel sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 217–237, Jan. 2013.
[21] 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.
[22] G. Moser and S. B. Serpico, "Combining support vector machines and Markov random fields in an integrated framework for contextual image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2734–2752, May 2013.
[23] B. Zhang, S. Li, X. Jia, L. Gao, and M. Peng, "Adaptive Markov random field approach for classification of hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 5, pp. 973–977, Sep. 2011.
[24] J. Li, J. M. 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.
[25] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral–spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 809–823, Mar. 2012.
[26] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral–spatial classification of hyperspectral data using loopy belief propagation and active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 844–856, Feb. 2013.
[27] Q. Jackson and D. A. Landgrebe, "Adaptive Bayesian contextual classification based on Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2454–2463, Nov. 2002.
[28] F. Melgani and S. B. Serpico, "A Markov random field approach to spatio-temporal contextual image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 11, pp. 2478–2487, Nov. 2003.
[29] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4085–4098, Nov. 2010.
[30] L. Sun, Z. Wu, Z. Wei, J. Liu, and X. Li, "Supervised hyperspectral image classification combining sparse unmixing and spatial constraint," in *Proc. Int. CVRS*, Dec. 2012, pp. 110–115.
[31] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984.
[32] O. Eches, N. Dobigeon, and J. Y. Tourneret, "Enhancing hyperspectral image unmixing with spatial correlations," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4239–4247, Nov. 2011.
[33] B. D. Ripley, *Statistical Inference for Spatial Processes*. Cambridge, U.K.: Cambridge Univ. Press, 1991.
[34] 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.
[35] T. Goldstein and S. Osher, "The split Bregman method for L1-regularized problems," *SIAM J. Imaging Sci.*, vol. 2, no. 2, pp. 323–343, Apr. 2009.
[36] C. Wu and X. C. Tai, "Augmented Lagrangian method, dual methods, split Bregman iteration for ROF, vectorial TV, high order models," *SIAM J. Imaging Sci.*, vol. 3, no. 3, pp. 300–339, Jul. 2010.
[37] E. Esser, "Applications of Lagrangian-Based Alternating Direction Methods and Connections to Split Bregman," University California Los Angeles, Los Angeles, CA, USA, CAM Tech. Rep. 09-31, 2009.
[38] M. Schmidt and K. Alahari, "Generalized fast approximate energy minimization via graph cuts: Alpha-expansion beta-shrink moves," *Arxiv Preprint arXiv:1108.5710*, 2011.
[39] M. D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4484–4502, Nov. 2012.
[40] J. Li, P. R. Marpu, A. Plaza, J. M. Bioucas-Dias, and J. A. Benediktsson, "Generalized composite kernel framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4816–4829, Sep. 2013.
[41] J. L Marroquin, E. A. Santana, and S. Botello, "Hidden Markov measure field models for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 11, pp. 1380–1387, Nov. 2003.
[42] M. A. T. Figueiredo, "Bayesian image segmentation using Gaussian field priors," in *Proc. Energy Minimization Methods Comput. Vis. Pattern Recog.*, St. Augustine, FL, USA, Nov. 2005, pp. 74–89.
[43] M. A. T. Figueiredo, "Bayesian image segmentation using wavelet-based priors," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, St. Augustine, FL, USA, Nov. 2005, pp. 437–443.
[44] J. Lellmann, J. Kappes, J. Yuan, F. Becker, and C. Schnorr, "Convex multi-class image labeling by simplex-constrained total variation," in *Proc. 2nd Int. Conf. Scale Space Variational Methods Comput. Vis.*, Voss, Norway, Jun. 2009, pp. 150–162.
[45] Q. Q. Yuan, L. P. Zhang, and H. F. Shen, "Hyperspectral image denoising employing a spectral–spatial adaptive total variation model," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3660–3677, Oct. 2012.
[46] G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. New York, NY, USA: Wiley, 1997.

Le Sun (S'13) was born in Jiangsu, China, in 1987. He received the B.S. degree from the School of Science, Nanjing University of Science and Technology, Nanjing, China, in 2009, where he is currently working toward the Ph.D. degree in the School of Computer Science and Engineering.

His research interests are in the area of spectral unmixing, hyperspectral classification, image processing, sparse representation, and compressive sensing.

Zebin Wu (M'13) was born in Zhejiang, China, in 1981. He received the B.S. and Ph.D. degrees from the School of Computer Science and Engineering, Nanjing University of Science and Technology (NJUST), Nanjing, China, in 2003 and 2008, respectively.

He is currently an Associate Professor with the School of Computer Science and Engineering, NJUST. His research interests include virtual reality and system simulation, remote-sensing information processing, and distributed computing.



Jianjun Liu was born in Jiangsu, China, in 1987. He received the B.S. degree from the School of Science, Nanjing University of Science and Technology, Nanjing, China, in 2009, where he is currently working toward the Ph.D. degree in the School of Computer Science and Engineering.

His research interests are in the area of spectral unmixing, hyperspectral classification, image processing, sparse representation, and compressive sensing.

Liang Xiao was born in Hunan, China, in 1976. He received the B.S. degree in applied mathematics and the Ph.D. degree in computer science from the Nanjing University of Science and Technology (NJUST), Nanjing, China, in 1999 and 2004, respectively.

From 2006 to 2008, he was a Postdoctoral Research Fellow with the Pattern Recognition Laboratory, NJUST. From 2009 to 2010, he was a Postdoctoral Fellow with Rensselaer Polytechnic Institute, Troy, NY, USA. He is currently a Professor with the School of Computer Science, NJUST. His main research areas include inverse problems in image processing, scientific computing, data mining, and pattern recognition.

Zhihui Wei was born in Jiangsu, China, in 1963. He received the B.S., M.S., and Ph.D. degrees from South East University, Nanjing, China, in 1983, 1986, and 2003, respectively.

He is currently a Professor and Doctoral Supervisor with Nanjing University of Science and Technology (NJUST), Nanjing. His main research areas include partial differential equations, mathematical image processing, image modeling, multiscale analysis, video and image coding, sparse representation, and compressive sensing. His current research is focused on the theory of image sampling, multiscale geometrical analysis, sparse representation, and compressive sensing.