Semisupervised Sparse Subspace Clustering Method With a Joint Sparsity Constraint for Hyperspectral Remote Sensing Images

Shaoguang Huang, *Student Member, IEEE*, Hongyan Zhang, *Senior Member, IEEE*, and Aleksandra Pižurica, *Senior Member, IEEE*

***Abstract*—Sparse subspace clustering (SSC), as an effective subspace clustering technique, has been widely applied in the remote sensing community, demonstrating a superior performance over the traditional methods such as k-means. In this paper, we propose a unified framework for hyperspectral image (HSI) clustering, which incorporates spatial information and label information in an SSC model, aiming at generating a more precise similarity matrix. The spatial information is included through a joint sparsity constraint on the coefficient matrix of each local region. Pixels within a local region are encouraged to select a common set of samples in the subspace-sparse representation, which greatly promotes the connectivity of the similarity matrix. We incorporate the available label information effectively within the same framework, by zeroing the entries of the sparse coefficient matrix, which correspond to the data points from different classes. An optimization algorithm is derived based on the alternating direction method of multipliers for the resulting model. Experimental results on real HSIs demonstrate a superior performance over the related state-of-the-art methods.**

***Index Terms*—Hyperspectral image (HSI), joint sparsity, semisupervised clustering, sparse subspace clustering (SSC).**

I. Introduction

HYPERSPECTRAL images (HSIs), captured with hundreds of spectral bands, offer richer information about the imaged objects than the multispectral images. Therefore, HSIs have become a powerful and valuable tool for various applications, including astronomy, geosciences, surveillance, defense and security [1], [2], agriculture [3], [4], and environmental monitoring [5]–[7]. HSI clustering is usually regarded as a fundamental step in these applications, aiming at assigning the pixels into different groups in an unsupervised way, where the pixels in the same group are more similar to each other than to those in different groups. However, due to the large spectral variability, noise and complex structure in the image of a scene, clustering of HSIs is a very challenging task [8].

Traditional clustering methods such as k-means [9], fuzzy c-means (FCM) [10], and generalized principal component analysis [11] have been widely used in remote sensing. However, directly applying such methods on HSIs often produces the clustering maps with a large amount of impulse noise, due to the limited discriminative information in spectral domain, complexity of ground objects, and large diversity of spectral signatures in the same class [12].

In recent years, sparse subspace clustering (SSC) [13] has emerged as an effective method for HSI clustering, providing the current state-of-the-art performance [8], [14]–[18]. SSC relies on a self-representation model where the input matrix is employed as the dictionary. The method postulates that each sample in a union of subspaces can be represented by a linear combination of other samples in the data set and all the involved samples in this representation are in the same subspace as the test sample. This indicates that given the dictionary for an input sample in the subspace $\mathcal{S}_l$, there exist a sparse representation vector whose nonzero entries correspond to the samples in the same subspace $\mathcal{S}_l$ [13], which is called subspace-sparse representation. As the subspace-sparse representation of the input matrix reflects the memberships of each sample, it is employed to build a similarity matrix, which is further applied within the spectral clustering framework [19]. The construction of the similarity matrix is a crucial step for the SSC model. In this paper, we will mainly focus on the generation of a well-designed similarity matrix.

The performance of SSC is limited by the fact that it treats each pixel independently without considering the spatial distribution of data points during the subspace-sparse representation, which makes the sparse coefficients sensitive to noise. Various extensions of the SSC model have been proposed to exploit spatial information in HSIs and achieved a significant improvement over SSC [8], [15], [16], [20]. The approach of [8] imposes a smoothness constraint on the neighboring sparse

Manuscript received July 27, 2018; revised November 13, 2018 and December 12, 2018; accepted January 14, 2019. Date of publication February 17, 2019; date of current version March 25, 2019. This work was supported in part by the Fonds voor Wetenschappelijk Onderzoek project: G.OA26.17N, in part by Grants from the China Scholarship Council (CSC) and UGent Bijzonder Onderzoeksfonds cofunding-CSC, and in part by the National Natural Science Foundation of China under Grants 61871298 and 41711530709. *(Corresponding author: Shaoguang Huang.)*

S. Huang and A. Pižurica are with the Department of Telecommunications and Information Processing, IPI, IMEC Research Group, Ghent University, 9000 Ghent, Belgium (e-mail: Shaoguang.Huang@Ugent.be; Aleksandra.Pizurica@UGent.be).

H. Zhang is with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China (e-mail: zhanghongyan@whu.edu.cn).

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/JSTARS.2019.2895508

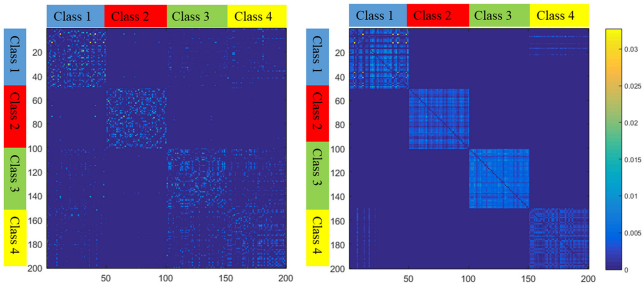

Fig. 1. Similarity matrix of four classes obtained by SSC (left) and the proposed JSSC-L (right) with 1% labeled samples in the *Indian Pines* image.

coefficients within square patches. Another ℓ_2-norm based spatial regularizer acting on the coefficients of horizontally and vertically adjacent pixels was proposed in [15] to promote piecewise smoothness of sparse coefficients. An object-oriented SSC method was proposed in [20], which extracts spectral-spatial features in each segmented object by a reweighted mass center learning method. A kernel version of SSC incorporating a max pooling of the sparse coefficient matrix was presented in [16].

However, none of the modified SSC models mentioned above incorporates label information in the construction of similarity matrix. Hence, their subspace-sparse representation is unsupervised, leaving room for improving the performance when labeled data are available. In order to utilize the available label information, a semisupervised SSC model was recently proposed in [21], where the class probability matrix of unlabeled samples was first calculated by using the labeled samples and then the whole probability matrix of all samples was applied as a weight matrix on the sparse matrix with the purpose of improving the global similarity structure among samples. However, spatial information was not taken into account and thus impulse noise was degrading significantly the resulted map.

To overcome the above limitations, in this paper, we propose a semisupervised SSC model in combination with spatial information of HSI, with the purpose of building a more accurate similarity matrix. We refer to the proposed method as joint SSC with label information (JSSC-L) throughout this paper. Fig. 1 shows the similarity matrix produced by SSC and JSSC-L where 1% labeled samples are utilized. There are 4 clusters each consisting of 50 samples from the *Indian Pines* image, and the pixels are sequentially arranged from 4 classes as shown in Fig. 1. In [13], it was explained that the ideal similarity matrix should be block-diagonal. Fig. 1 reveals that our proposed method preserves such structure much better than SSC.

The benefits of the JSSC-L model are mainly due to the following two aspects. First, unlike the ℓ_1 norm constraints on the coefficient matrix in the aforementioned approaches [8], [15], [16], [20], [21], we exploit the spatial information of HSI by enforcing a joint sparsity constraint as $\ell_{1,2}$ norm, on each local region in the subspace-sparse representation. The local regions are typically homogeneous in HSI [22]–[29] and we obtain these local regions by a superpixel segmentation technique, yielding nonoverlapping superpixels with an adaptive shape and size. Each superpixel is typically comprised of the highly similar materials, and consequently the spectral signatures within one superpixel are mutually similar. To preserve such a relationship, we integrate a joint sparsity constraint for the coefficients of pixels in each superpixel with the SSC model, promoting this way the graph connectivity. The second aspect is that we utilize the label information in HSI to build a more precise connection between the data points. As neighboring pixels in a local region are mostly in the same class, we first propagate the label information to each superpixel by exploiting such spatial information, which allows us to employ the label information as much as possible. Then, the links between every pair of data points are refined in the way where we set the sparse coefficient to zero if the two corresponding data points are from different classes. Therefore, the JSSC-L model is able to capture the global data structure by respecting the label information, which ensures that the generated similarity matrix follows well the genuine block-diagonal structure. Note that the way we employ the label information is much different from the label usage in supervised classification methods like sparse representation classification [30] and the block SSC [31], where class labels are employed to build a structured dictionary. To solve the resulting optimization problem, an iterative solver based on alternating direction method of multipliers (ADMM) is derived. Experimental results on three real data sets demonstrate the effectiveness of the proposed method with a significant performance improvement over the existing methods in terms of overall accuracy (OA).

The preliminary work was reported in a conference [32]. In comparison with the conference version, which only exploits the spatial information with a joint sparsity constraint on local regions, here we extend the model further to a semisupervised SSC model by utilizing label information, which yields a more precise similarity matrix. The major contributions of the paper can be summarized as follows.

1) A unified framework is proposed for HSI clustering, which incorporates spatial information and label information, leading to the significantly improved clustering performance.
2) The spatial information of HSI is included through a joint sparsity constraint on the coefficient matrices of local regions to promote the graph connectivity.
3) The label information of HSI is utilized to refine the estimation of the similarity matrix, which excludes the connections of pixels in different classes directly by setting the corresponding coefficients to zeros. The resulted similarity matrix is much closer to the block-diagonal structure. This method is general and can be easily applied to other existing SSC-based models.
4) An efficient ADMM-based algorithm is derived to solve the resulting optimization problem.

The rest of this paper is organized as follows. Section II briefly introduces the clustering of HSIs with the SSC model. Section III describes the proposed JSSC-L model and the resulting optimization problem. Section IV presents the experimental results on real data and Section V concludes the paper.

Algorithm 1: HSI Clustering with the SSC Model.

1: **Input**: An input matrix $\mathbf{Y} \in \mathbb{R}^{B\times MN}$, λ and the number of clusters c.
2: Calculate the sparse coefficient matrix $\mathbf{C}$ by (1).
3: Normalize the columns of $\mathbf{C}$ by $\mathbf{c}_i = \mathbf{c}_i/\|\mathbf{c}_i\|_\infty$.
4: Compute the similarity matrix $\mathbf{W}$ by (2).
5: Obtain the clustering results by applying $\mathbf{W}$ in the spectral clustering framework [19].
6: **Output**: A clustering map.

II. HSI Clustering With the SSC Model

The spectral signature of each pixel in HSI is a vector with each entry corresponding to the spectral reflection value of object in a specific wavelength. A three-dimensional HSI data cube with the size of $M \times N \times B$, where M and N represent the height and the width of the data, respectively, and B the number of bands, is first flattened to a two-dimensional matrix $\mathbf{Y} \in \mathbb{R}^{B\times MN}$ with each column being a spectral signature. Under the assumption that each pixel in a union of subspaces can be sparsely represented as a linear or affine combination of the others from the same subspace [13], [33], the sparse coefficient matrix $\mathbf{C} \in \mathbb{R}^{MN\times MN}$ of $\mathbf{Y}$ can be derived by solving the following optimization problem:

$$\begin{aligned} \arg\min_{\mathbf{C}} \ & \|\mathbf{C}\|_1 + \frac{\lambda}{2}\|\mathbf{Y}-\mathbf{Y}\mathbf{C}\|_F^2 \\ \text{s.t. } & \operatorname{diag}(\mathbf{C}) = \mathbf{0},\ \mathbf{1}^T\mathbf{C} = \mathbf{1}^T \end{aligned} \tag{1}$$

where $\|\mathbf{C}\|_1 = \sum_i \sum_j |C_{ij}|$; $\mathbf{1}^T$ is a row vector of size $1 \times MN$; $\operatorname{diag}(\mathbf{C})$ is a diagonal matrix whose entries outside the main diagonal are zero and λ is a parameter, which controls the balance between the data fidelity and the sparsity of the coefficient matrix. The first constraint $\operatorname{diag}(\mathbf{C}) = \mathbf{0}$ is introduced to avoid the trivial solution of representing a sample by itself and the second constraint $\mathbf{1}^T\mathbf{C} = \mathbf{1}^T$ indicates the case of affine subspace.

The model in (1) can be solved by ADMM [34]. As the nonzero entries of $\mathbf{C}$ indicate which data points will be selected in the subspace-sparse representation of $\mathbf{Y}$, the matrix $\mathbf{C}$ indicates the connections between each data point of $\mathbf{Y}$. The similarity matrix $\mathbf{W} \in \mathbb{R}^{MN\times MN}$ is given by

$$\mathbf{W} = |\mathbf{C}| + |\mathbf{C}|^T. \tag{2}$$

The symmetric structure of $\mathbf{W}$ makes sure that each pair of samples are connected to each other if either side is selected to represent another, which results in a strengthened connection of the graph. With spectral graph theory, the clustering results can be achieved by applying k-means to a subset of eigenvectors of the Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{W}$ where $\mathbf{D} \in \mathbb{R}^{MN\times MN}$ is a diagonal matrix with $D_{ii} = \sum_j W_{ij}$ [19]. The SSC model is summarized in Algorithm 1.

III. Proposed JSSC-L Model for HSIs

In this section, we first introduce a joint sparsity based SSC (JSSC) model, which only incorporates the spatial information of HSIs. Then, our complete JSSC-L model employing the label information is presented. Finally, we develop an optimization algorithm for the resulting model based on ADMM.

A. JSSC Model

In the subspace-sparse representation of SSC, the sparse coefficient vector of each pixel of $\mathbf{Y}$ is calculated individually and independently. Various factors, such as limited discriminative information in spectral domain, noise, and large diversity of spectral signatures belonging to the same class in HSIs [8], degrade the construction of the similarity matrix, deteriorating thereby spectral clustering performance. In practice, pixel values are spatially correlated, which means that pixels within a local region belong to the same class with high probability [23]–[25], [27], [29].

Here, we incorporate the spatial information with a joint sparsity constraint to preserve the dependencies between HSI pixels in each local region.

Suppose that an HSI is segmented into p nonoverlapping superpixels [35]. Each superpixel is regarded as a homogeneous region belonging to a particular class. We assume that each pixel within the same superpixel selects the same set of pixels in the data representation but with different coefficients. The resulting JSSC model can be described as follows:

$$\begin{aligned} \arg\min_{\mathbf{C}} \ & \sum_{i=1}^{p} w_i\|\mathbf{C}_i\|_{1,2} + \frac{\lambda}{2}\|\mathbf{Y}-\mathbf{Y}\mathbf{C}\|_F^2 \\ \text{s.t. } & \operatorname{diag}(\mathbf{C}) = \mathbf{0},\ \mathbf{1}^T\mathbf{C} = \mathbf{1}^T \end{aligned} \tag{3}$$

where $\mathbf{C}_i \in \mathbb{R}^{MN\times n_i}$ is a sparse matrix that corresponds to the n_i pixels in ith superpixel; $\|\mathbf{C}_i\|_{1,2}$ is the $\ell_{1,2}$ norm defined as $\sum_{j=1}^{MN} \|\mathbf{c}_i^j\|_2$. Here $\mathbf{c}_i^j$ is the jth row of $\mathbf{C}_i$ and w_i is a normalized weight defined as

$$w_i = \frac{\sqrt{n_i/p}}{\sum_{i=1}^{p}\sqrt{n_i/p}} \tag{4}$$

to balance the joint sparsity constraint for each superpixel. The first term with $\ell_{1,2}$ norm takes care that the pixels within one superpixel are likely to select a common set of samples in the data representation, promoting this way the graph connectivity within the local region.

B. JSSC-L Model

In order to better preserve the original data structure, we integrate the label information of the observed pixels explicitly into the JSSC model, resulting in its semisupervised extension JSSC-L. The actual structure of ideal similarity matrix is block-diagonal [13], which means that there are no connections between pixels from different classes. To comply with such diagonal structure, we set the coefficients $\mathbf{C}_{i,j}$ to zero if pixels i and j are from different classes. Since local regions of HSI are typically homogeneous (belonging to the same class), we first propagate the label information to each superpixel by a majority voting strategy. If a superpixel contains labeled samples from different classes, the class to which most of them belong will

be assigned. The resulting optimization problem with respect to $\mathbf{C}$ is

$$\arg\min_{\mathbf{C}} \sum_{i=1}^{p} w_i \|\mathbf{C}_i\|_{1,2} + \frac{\lambda}{2}\|\mathbf{Y}-\mathbf{YC}\|_F^2 \quad \text{s.t. } \operatorname{diag}(\mathbf{C})=\mathbf{0},\ \mathbf{1}^T\mathbf{C}=\mathbf{1}^T,\ \mathcal{P}_{\mathbb{L}}(\mathbf{C})=\mathbf{0} \tag{5}$$

where $\mathbb{L}$ is a set of pairs $\{i, j\}$ such that i and j are labeled pixels from different classes, and $\mathcal{P}_{\mathbb{L}}(\mathbf{C})$ is a projection operator that extracts the entries in $\mathbf{C}$ whose indices are in $\mathbb{L}$. Obviously, the third constraint $\mathcal{P}_{\mathbb{L}}(\mathbf{C})=\mathbf{0}$ makes use of the label information by setting the entries for pixels pairs from different classes to zero.

We merge the constraints $\operatorname{diag}(\mathbf{C})=0$ and $\mathcal{P}_{\mathbb{L}}(\mathbf{C})=\mathbf{0}$ into $\mathcal{P}_{\mathbb{G}}(\mathbf{C})=\mathbf{0}$, where $\mathbb{G}$ is the union of set $\mathbb{L}$ and set $\{i, i\}$. The objective function (5) can be reformulated as

$$\arg\min_{\mathbf{C}} \sum_{i=1}^{p} w_i \|\mathbf{C}_i\|_{1,2} + \frac{\lambda}{2}\|\mathbf{Y}-\mathbf{YC}\|_F^2 \quad \text{s.t. } \mathbf{1}^T\mathbf{C}=\mathbf{1}^T,\ \mathcal{P}_{\mathbb{G}}(\mathbf{C})=\mathbf{0}. \tag{6}$$

The optimization problem (6) will be solved by an ADMM-based method as described in the next section.

Once we obtain the sparse matrix $\mathbf{C}$, the similarity matrix is calculated using (2) and further applied within the spectral clustering method to yield the clustering results in the same way as in SSC.

C. Optimization Method

In this section, an efficient optimization algorithm is derived. We base our approach on the general algorithm for solving convex problems. ADMM [34] splits a complicated optimization problem into smaller problems that are easier solved and often has closed-form solutions. The subproblems are iteratively solved until stop criterion is satisfied.

We first reformulate (6) as the following optimization problem:

$$\arg\min_{\mathbf{C},\mathbf{A}} \sum_{i=1}^{p} w_i \|\mathbf{A_i}\|_{1,2} + \frac{\lambda}{2}\|\mathbf{Y}-\mathbf{YC}\|_F^2 \quad \text{s.t. } \mathbf{1}^T\mathbf{C}=\mathbf{1}^T,\ \mathbf{C}=\mathbf{A}-\mathcal{P}_{\mathbb{G}}(\mathbf{A}) \tag{7}$$

where $\mathbf{A}\in\mathbb{R}^{MN\times MN}$ is an auxiliary variable.

We denote the resulting augmented Lagrangian function by $f_\mu(\mathbf{C},\mathbf{A};\mathbf{Y_1},\mathbf{Y_2})$ as follows:

$$\sum_{i=1}^{p} w_i\|\mathbf{A_i}\|_{1,2} + \frac{\lambda}{2}\|\mathbf{Y}-\mathbf{YC}\|_F^2 + \langle \mathbf{Y_1}, \mathbf{1}^T\mathbf{C}-\mathbf{1}^T\rangle + \langle \mathbf{Y_2}, \mathbf{C}-\mathbf{A}+\mathcal{P}_{\mathbb{G}}(\mathbf{A})\rangle + \frac{\mu}{2}(\|\mathbf{1}^T\mathbf{C}-\mathbf{1}^T\|_2^2 + \|\mathbf{C}-\mathbf{A}+\mathcal{P}_{\mathbb{G}}(\mathbf{A})\|_F^2) \tag{8}$$

where $\mathbf{Y}_1\in\mathbb{R}^{1\times MN}$, $\mathbf{Y}_2\in\mathbb{R}^{MN\times MN}$, and μ are the penalty parameters for the appended terms and $\langle \mathbf{V}_1,\mathbf{V}_2\rangle$ denotes the trace of $\mathbf{V}_1^T\mathbf{V}_2$. Due to the separable structure of f_μ, we can solve each of $\mathbf{C}$ and $\mathbf{A}$ separately by fixing one when solving another. The concrete steps are described next.

1) Update C: The objective function with respect to $\mathbf{C}$ is given by

$$\mathbf{C}^{k+1} = \arg\min_{\mathbf{C}} \frac{\lambda}{2}\|\mathbf{Y}-\mathbf{YC}\|_F^2 + \frac{\mu^k}{2}\left(\left\|\mathbf{1}^T\mathbf{C}-\mathbf{1}^T + \frac{\mathbf{Y}_1^k}{\mu^k}\right\|_2^2 + \left\|\mathbf{C}-\mathbf{A}^k+\mathcal{P}_{\mathbb{G}}(\mathbf{A}^k)+\frac{\mathbf{Y}_2^k}{\mu^k}\right\|_F^2\right). \tag{9}$$

By setting the first-order derivative to zero, a closed-form solution is obtained as

$$\mathbf{C}^{k+1} = (\lambda\mathbf{Y}^T\mathbf{Y}+\mu^k(\mathbf{I}+1))^{-1}(\lambda\mathbf{Y}^T\mathbf{Y} + \mu^k(\mathbf{A}^k-\mathcal{P}_{\mathbb{G}}(\mathbf{A}^k)+1) - \mathbf{1}\mathbf{Y}_1^k - \mathbf{Y}_2^k). \tag{10}$$

2) Update A: The objective function with respect to $\mathbf{A}$ is

$$\mathbf{A}^{k+1} = \arg\min_{\mathbf{A}} \sum_{i=1}^{p} w_i\|\mathbf{A}_i\|_{1,2} + \frac{\mu^k}{2}\left\|\mathbf{C}^{k+1}-\mathbf{A} + \mathcal{P}_{\mathbb{G}}(\mathbf{A}) + \frac{\mathbf{Y}_2^k}{\mu^k}\right\|_F^2 \tag{11}$$

which can be separated into p subproblems corresponding to each superpixel

$$\arg\min_{\mathbf{A}_i} w_i\|\mathbf{A}_i\|_{1,2} + \frac{\mu^k}{2}\left\|\mathbf{C}_i^{k+1}-\mathbf{A}_i+\mathcal{P}_{\mathbb{G}}(\mathbf{A}_i)+\frac{\mathbf{Y}_{2_i}^k}{\mu^k}\right\|_F^2 \tag{12}$$

where $\mathbf{Y}_{2_i}\in\mathbb{R}^{MN\times n_i}$ is a matrix with each column indexed by the pixels in ith superpixel. Let $\mathbf{a}_{i,t}^{k+1}$, $\mathbf{c}_{i,t}^{k+1}$, and $\mathbf{Y}_{2_i,t}^k$ be the rows of matrices $\mathbf{A}_i^{k+1}$, $\mathbf{C}_i^{k+1}$, and $\mathbf{Y}_{2_i}^k$, respectively. We solve each subproblem for $t=1,2,\ldots,n_i$ as follows:

$$\mathbf{a}_{i,t}^{k+1} = \arg\min_{\mathbf{a}} w_i\|\mathbf{a}\|_2 + \frac{\mu^k}{2}\|\mathbf{a}-\mathcal{P}_{\mathbb{G}}(\mathbf{a})-\mathbf{z}\|_2^2 \tag{13}$$

where $\mathbf{z}=\mathbf{c}_{i,t}^{k+1}+\mathbf{Y}_{2_i,t}^k/\mu^k$. The problem (13) can be solved using

$$\mathbf{a}_{i,t}^{k+1} = \tilde{\mathbf{a}} - \mathcal{P}_{\mathbb{G}}(\tilde{\mathbf{a}}) \tag{14}$$

where $\tilde{\mathbf{a}}=(1-w_i/\mu^k/\|\mathbf{z}\|_2)_+\mathbf{z}$ and $(\mathbf{x})_+$ is a vector with entries $\max(x_i,0)$.

3) Update Other Parameters: The next step is to update the multipliers $\mathbf{Y}_1,\mathbf{Y}_2$, and μ, which can be achieved by

$$\mathbf{Y}_1^{k+1} = \mathbf{Y}_1^k + \mu^k(\mathbf{1}^T\mathbf{C}^{k+1}-\mathbf{1}^T) \tag{15}$$

$$\mathbf{Y}_2^{k+1} = \mathbf{Y}_2^k + \mu^k(\mathbf{C}^{k+1}-\mathbf{A}^{k+1}+\mathcal{P}_{\mathbb{G}}(\mathbf{A}^{k+1})) \tag{16}$$

$$\mu^{k+1} = \rho\mu^k \tag{17}$$

where $\rho\geq 1$ is a parameter that controls the convergence speed of the optimization. The three updating steps are executed iteratively until the stopping criterion is satisfied, i.e., $\|\mathbf{1}^T\mathbf{C}^{k+1}-\mathbf{1}^T\|_\infty<\varepsilon$ and $\|\mathbf{C}^{k+1}-\mathbf{A}^{k+1}\|_\infty<\varepsilon$, or $k>$ MaxIter, where MaxIter is the maximum number of iteration. We summarize the optimization algorithm of JSSC-L in Algorithm 2.

Algorithm 2: ADMM for Solving JSSC-L Model.

Input: An input matrix $\mathbf{Y} \in \mathbb{R}^{B\times MN}$, λ, the number of super-pixel p and indice set $\mathbb{G}$.
Initialize: $\mathbf{A}^0 = \mathbf{0}$, $\mathbf{Y}_1^0 = \mathbf{0}$, $\mathbf{Y}_2^0 = \mathbf{0}$, $\mu^0 = 1$, $\rho = 1.1$, $\varepsilon = 10^{-5}$, $MaxIter$, $k \leftarrow 0$.
Do
  Update $\mathbf{C}$ by (10).
  Update $\mathbf{A}$ by (11).
  Update $\mathbf{Y}_1$ by (15).
  Update $\mathbf{Y}_2$ by (16).
  Update μ by (17).
While ($\|\mathbf{1}^T\mathbf{C}^{k+1} - \mathbf{1}^T\|_\infty > \varepsilon$ or $\|\mathbf{C}^{k+1} - \mathbf{A}^{k+1}\|_\infty > \varepsilon$ and $k < MaxIter$)
Output: Coefficient matrix $\mathbf{C}$.

IV. Experimental Results and Analysis

Experiments are conducted on following three real HSIs to validate the performance of the proposed method: 1) *Indian Pines* image; 2) *Pavia University* image; and 3) *Salinas* image. The results of two widely used clustering methods FCM [10] and k-means [9], the clustering by fast search and find of density peaks (CFSFDP) [36] and state-of-the-art methods SSC [13], L2-SSC [15], and a class probability propagation of supervised information based on SSC (CPPSSC) [21] are reported for comparison. In addition, we also report the performance of our method JSSC, which does not utilize label information as reference. We used the online source codes of SSC[1] and CFSFDP[2] in our experiments reported here.

Two common performance measures: OA and Kappa coefficient (κ) are used for quantitative assessment of the clustering performances. The optimal parameters of SSC and L2-SSC are acquired as in [8], [15], and [16]. For the methods JSSC, CPPSSC, and JSSC-L, we manually tune the parameters by a grid-search strategy. JSSC shares the same segmentation map of JSSC-L. Note that JSSC-L reduces to JSSC when no label information is employed. In this paper, we randomly select 1% labeled samples for CPPSSC and JSSC-L, and the parameter λ of JSSC-L is set to 1 based on empirical optimization. To avoid a biased estimation, the experiments for CPPSSC and JSSC-L are repeated five times and the average results are reported.

A. Airborne/Visible Infrared Imaging Spectrometer (AVIRIS) Data Set: Indian Pines Image

We present here the experimental results on the *Indian Pines* image, which was acquired by the AVIRIS sensors from the North-western Indiana in June 1992. The image consists of 16 classes and has a size of $145 \times 145 \times 220$, with a spatial resolution of 20 m per pixel and 220 spectral reflectance bands in the wavelength range 0.4–2.5 μm. The spectral bands in 104–108, 150–163, and 200 are removed due to the water absorption. We select a subimage with the size of 85×70 as test data for computational efficiency, as it was done in [8]. There are four classes in the test data as shown in Table I. The false color image and ground truth are shown in Fig. 2(a) and (b).

The parameters for JSSC-L are set to $\lambda = 1$ and $p = 10$. The results reported in Table I and Fig. 2 indicate that our method JSSC-L achieves the best performance in terms of OA and κ. Compared with the two classical FCM and k-means methods and the CFSFDP method, most of the sparsity-based approaches show superior clustering accuracy, demonstrating the effectiveness of sparsity-based clustering methods in the present task. The JSSC-L yields a significant improvement over SSC and L2-SSC with OA enhancements of 29.84% and 27.17% respectively. Compared with CPPSSC, which also employs the label information, JSSC-L achieves a 15.6% OA improvement. The improvement over SSC is mainly due to the two aspects: 1) the use of spatial information (with a joint sparsity constraint on local regions), which yields 21.29% OA enhancement and 2) the employment of label information, which yields 8.55% improvement in terms of OA. In Fig. 2, it can be seen that the clustering map of the CPPSSC method, which only utilizes label information, is heavily affected by impulse noise. The clustering map of JSSC-L is much smoother, demonstrating the benefit of incorporating spatial information.

B. Reflective Optics System Imaging Spectrometer (ROSIS) Urban Data Set: Pavia University Image

The second experiment was conducted on an urban HSI: *Pavia University*. This image was collected by the ROSIS sensor during a flight campaign over Pavia, Northern Italy. The image size is 512×217. The geometric resolution is 1.3 m and 103 spectral bands are captured. In total, 20 water absorption bands in 108–112, 154–167, and 224 are removed in this experiment. A typical area with a size of 200×100 is extracted as the test data that includes eight classes in total. The false color image and ground truth can be found in Fig. 3(a) and (b).

We set $\lambda = 1$ and $p = 20$ as the optimal parameters for our method and report the clustering accuracy and maps in Table II and Fig. 3. JSSC-L yields again the best clustering performance in terms of OA and κ, with 33.85% and 23.75% improvement of OA over SSC and L2-SSC, respectively. All sparsity-based methods obtain better clustering results than FCM, k-means, and CFSFDP. In comparison with CPPSSC, our method also achieves 25.45% increase in OA. Table II also reveals that for the class 1 almost all the methods fail to obtain good results while our method reaches an accuracy of 100%. Moreover, Fig. 3 shows that the clustering map of our method is much closer to the ground truthfig.

C. AVIRIS Data Set: Salinas Image

The third experiment was conducted on the *Salinas* image, which was captured by the AVIRIS sensor over the Salinas Valley, CA, USA. The image size is $512 \times 217 \times 224$ with a high spatial resolution of 3.7 m per pixel. There are 16 classes in total. In total, 20 bands in 108–112, 154–167, and 224 are removed due to the water absorption. We select a typical region with a size of 100×80 as test data, consisting of six classes as

[1]http://www.ccs.neu.edu/home/eelhami/codes.htm
[2]https://people.sissa.it/∼ laio/Research/Res_clustering.php

TABLE I
CLUSTERING ACCURACY FOR *Indian Pines*

No.	Class name	FCM	k-means	CFSFDP	SSC	L2-SSC	CPPSSC	JSSC	JSSC-L
1	Corn-notill	62.39	69.85	28.46	60.00	61.09	71.62	74.03	**93.27**
2	Grass-trees	94.66	53.84	**100**	98.36	99.32	98.66	**100**	**100**
3	Soybean-notill	44.13	0	82.38	76.91	79.37	77.46	86.20	**99.59**
4	Soybean-mintill	63.83	57.59	50.73	50.68	54.89	76.78	87.79	**92.15**
	OA(%)	65.34	50.17	59.10	65.11	67.78	79.35	86.40	**94.95**
	κ	0.5118	0.2833	0.4409	0.5296	0.5629	0.7081	0.8069	**0.9286**

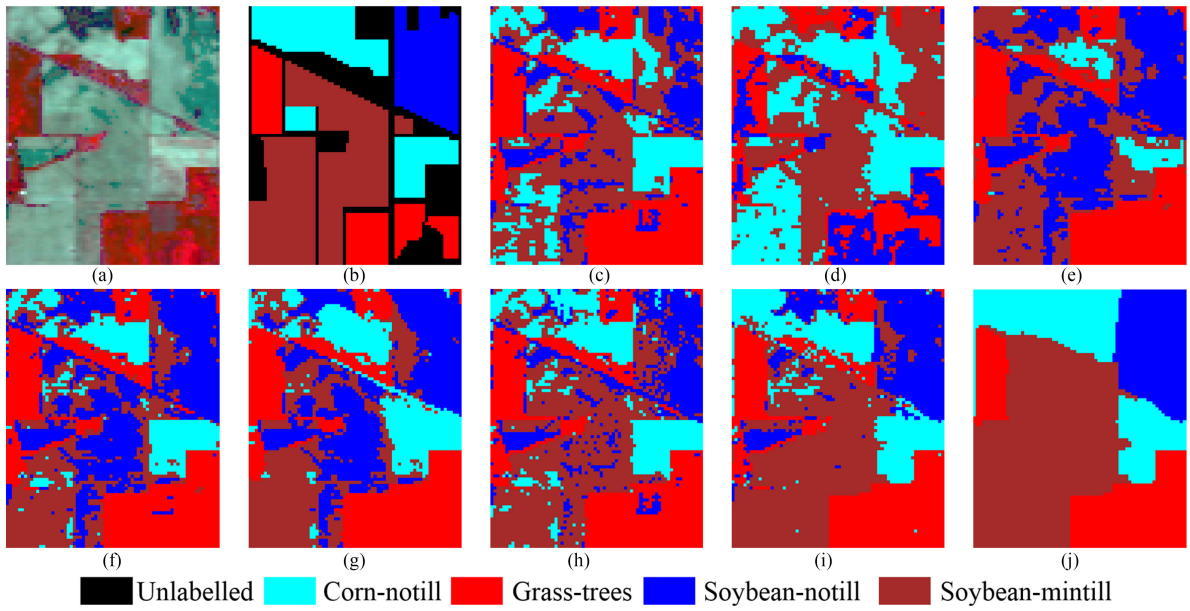

Fig. 2. *Indian Pines* image. (a) False color image, (b) ground truth, and clustering maps of (c) FCM, (d) k-means, (e) CFSFDP, (f) SSC, (g) L2-SSC, (h) CPPSSC (1% labeled samples), (i) JSSC, and (j) JSSC-L (1% labeled samples).

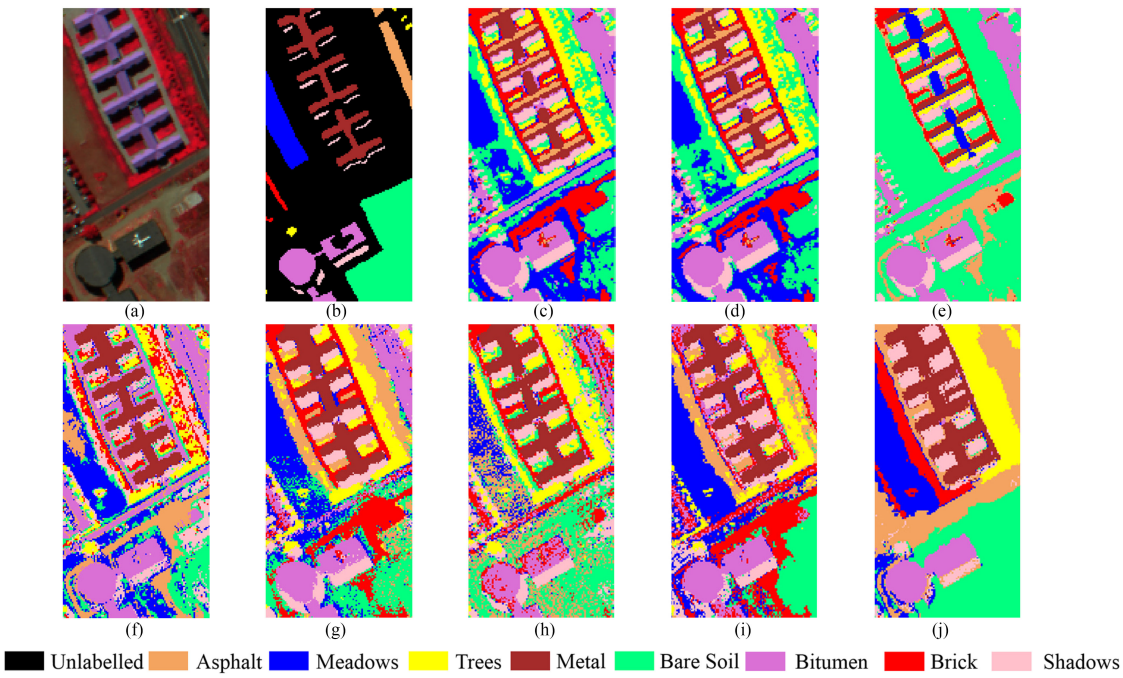

Fig. 3. *Pavia University* image. (a) False color image, (b) ground truth, and clustering maps of (c) FCM, (d) k-means, (e) CFSFDP, (f) SSC, (g) L2-SSC, (h) CPPSSC (1% labeled samples), (i) JSSC, and (j) JSSC-L (1% labeled samples).

TABLE II
CLUSTERING ACCURACY FOR *Pavia University*

No.	Class name	FCM	k-means	CFSFDP	SSC	L2-SSC	CPPSSC	JSSC	JSSC-L
1	Asphalt	0	0	0	6.12	0	26.82	0	**100**
2	Meadows	86.33	78.91	0	57.55	94.27	75.65	**99.84**	96.64
3	Trees	60.32	66.67	0	95.24	84.13	**100**	90.48	54.92
4	Metal	61.37	58.63	37.79	98.10	97.34	**99.42**	99.39	95.22
5	Bare Soil	40.72	44.43	74.76	39.12	51.78	55.90	63.42	**99.72**
6	Bitumen	**100**	**100**	**100**	98.49	97.21	82.07	94.07	88.14
7	Brick	1.06	0	0	0	39.36	53.19	51.06	**80.00**
8	Shadows	**100**	**100**	**100**	82.55	99.72	89.70	95.01	39.45
	OA(%)	58.53	58.60	56.34	61.51	71.61	70.99	76.82	**92.81**
	κ	0.5032	0.5	0.4310	0.5378	0.6554	0.6463	0.7142	**0.9047**

TABLE III
CLUSTERING ACCURACY FOR *Salinas*

No.	Class name	FCM	k-means	CFSFDP	SSC	L2-SSC	CPPSSC	JSSC	JSSC-L
1	Brocoli-1	99.01	**100**	**100**	99.21	100	98.84	97.14	**100**
2	Brocoli-2	42.48	0	0	61.50	99.69	99.97	**100**	**100**
3	Grapes-untrained	96.74	96.95	97.96	98.47	96.59	98.77	96.34	**99.67**
4	Lettuce-4wk	93.11	92.69	93.95	94.80	87.20	95.13	97.75	**100**
5	Lettuce-5wk	**100**	**100**	**100**	99.78	95.59	99.98	**100**	**100**
6	Lettuce-6wk	**100**	97.82	95.20	97.38	**100**	**100**	**100**	**100**
	OA(%)	90.95	86.02	86.44	93.93	96.35	98.71	97.87	**99.88**
	κ	0.8833	0.8189	0.8241	0.9217	0.9532	0.9834	0.9728	**0.9985**

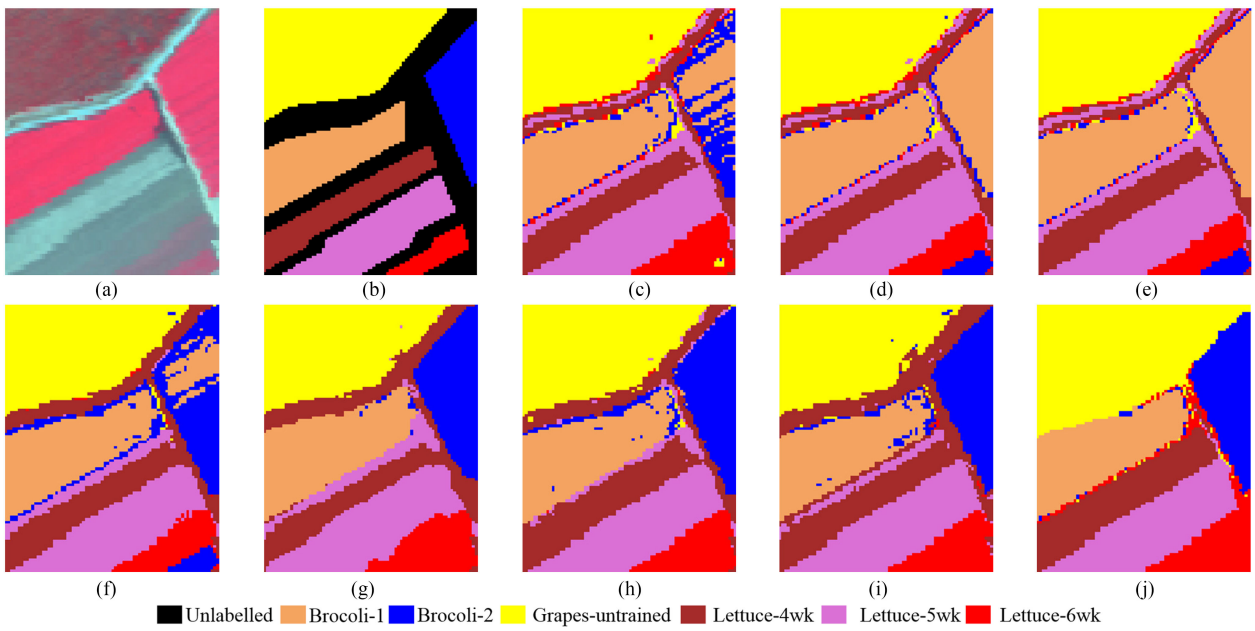

Fig. 4. *Salinas* image. (a) False color image, (b) ground truth, and clustering maps of (c) FCM, (d) k-means, (e) CFSFDP, (f) SSC, (g) L2-SSC, (h) CPPSSC (1% labeled samples), (i) JSSC, and (j) JSSC-L (1% labeled samples).

shown in Table III. False color image and ground truth are given in Fig. 4(a) and (b).

We set $\lambda = 1$ and $p = 10$ as the optimal parameters of JSSC-L, and report the clustering results for different methods in Table III and Fig. 4. It is clear that JSSC-L consistently yields a superior performance over others with an overall accuracy of 99.88%. For the classes 1, 2, 4, 5, and 6, the proposed method JSSC-L even achieves the clustering accuracy of 100%. In addition, when no label information is available, our method JSSC still yields a better result than other state-of-the-art approaches.

D. Analysis of Parameters

This section investigates the effect of parameters λ and p on the performance of JSSC-L model. We vary λ in the range of $\{10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}, 10^{3}\}$ and p in the range

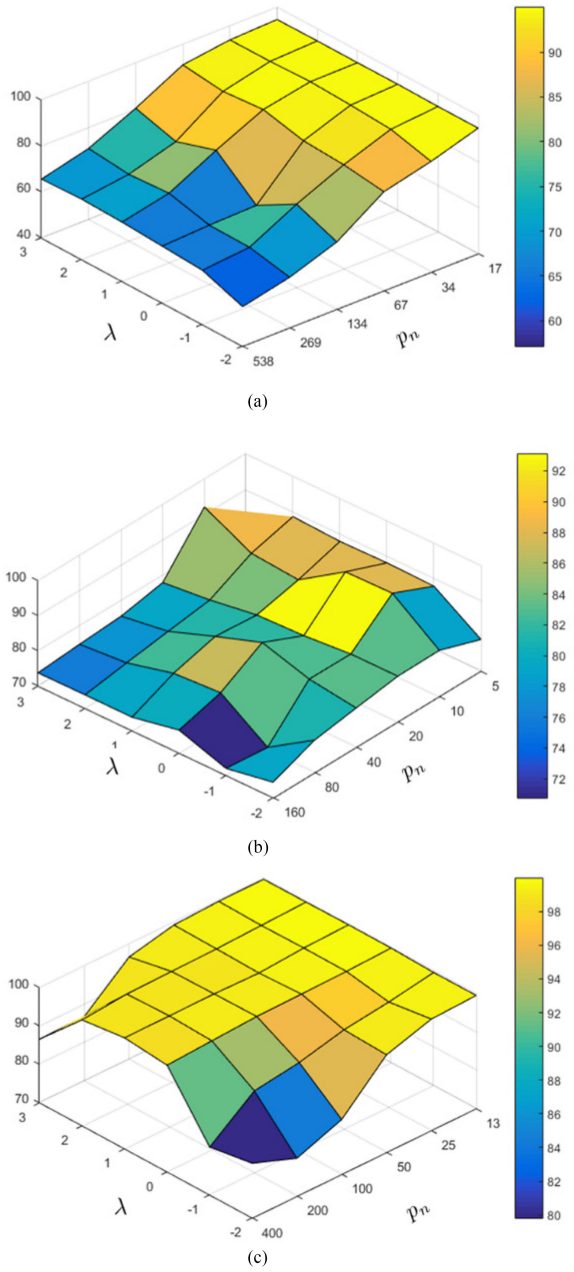

Fig. 5. Grid search of λ and p for JSSC-L in three data sets. (a) *Indian Pines*. (b) *Pavia University*. (c) *Salinas* (the λ-axis is in $\log_{10}$ and p_n is defined by $p/M/N \times 10^4$).

of $\{10, 20, 40, 80, 160, 320\}$ for the three data sets. The results are shown in Fig. 5. Note that the λ-axis is in $\log_{10}$ and p_n is a normalized parameter for p as defined by $p_n = p/M/N \times 10^4$. These results indicate that JSSC-L is able to achieve a good and rather stable performance for a wide range of these parameters, i.e., when $p_n \in [10, 67]$ and $\lambda \in [1, 100]$. Especially the performance is stable for the *Indian Pines* image and *Salinas* image. For the *Pavia University* image, even though the results of JSSC-L are less stable than for other two images, in most cases, the performance is still superior compared to SSC and L2-SSC. Based on the empirical analysis, we recommend the value of p to be set in the range of $[10, 67] \times M \times N/10^4$.

The results in Fig. 5 show that the number of superpixels p is a more important parameter compared with λ. When p is set properly, the performance is quite stable with varying λ. A larger value of p may capture the spatial information more precisely but it will degrade the spatial connections between some data points as the pixels are going to be treated more independently which limits the ability of joint sparsity constraint on promoting the graph connectivity, which is obviously demonstrated by a declining curve with a fixed λ in Fig. 5(a). A smaller value of p may result in merging some small homogeneous regions and neglecting this way some minor classes during the label propagation. If in addition some of these minor classes are not represented in the labeled data, a much too strong constraint may be imposed on the coefficient matrix by the available labeled data. This is the main reason why all the methods detect the grass-trees (unlabeled in the ground truth) in the top middle part of clustering maps in Fig. 2 except the JSSC-L model. An effective way to overcome such a degradation on the minor classes is to increase properly the number of superpixels to yield a more precise segmentation map. We report the clustering maps of JSSC-L under different values of p with 1% labeled samples in the *Indian Pines* image in Fig. 6.

The results show that the clustering accuracy is decreasing with the increasing values of p. The JSSC-L with $p = 10$ yields the best performance with a high accuracy of 96.90%, but in fact it does not completely match the false color image visually especially for the grass-trees on the top. The reason is that the ground truth is not labeled overall and the best result is reported according to the highest accuracy with respect to this ground truth. So the clustering map in the unlabeled area may be inaccurate. When p is set as 20 or 40, the clustering maps are much closer to the false color image than that of $p = 10$. Even though the accuracy is not the highest, visually the result in Fig. 6(e) can be regarded as the best one.

E. Generalization of the Existing Models

Compared with SSC and L2-SSC, the improved performance of JSSC-L is partly due to making use of label information. In fact, our efficient approach to exploit the label information can be readily incorporated into other off-the-shelf approaches generalizing them and improving their performance. Specifically, we add the constraint $\mathcal{P}_{\mathbb{G}}(\mathbf{C}) = 0$ in the SSC and L2-SSC model and refer to the resulting methods as SSC-L and L2-SSC-L, respectively. The results in Figs. 7 and 8 demonstrate that the clustering accuracy of both SSC and L2-SSC are significantly improved for all the tested data sets. For the *Pavia University* image, these improvements are nearly 20% and, for the *Indian Pines* image, the improvements are even larger: nearly 30%. Obviously, the estimation of the similarity matrix benefits from the label information, which can be effectively integrated into various method using our approach.

F. Effect of the Number of Labeled Samples

We also investigate the effect of the number of labeled samples on the performance of JSSC-L. The percentages of labeled samples per class are set in the range of $\{0, 1, 5, 10, 20\}$, and

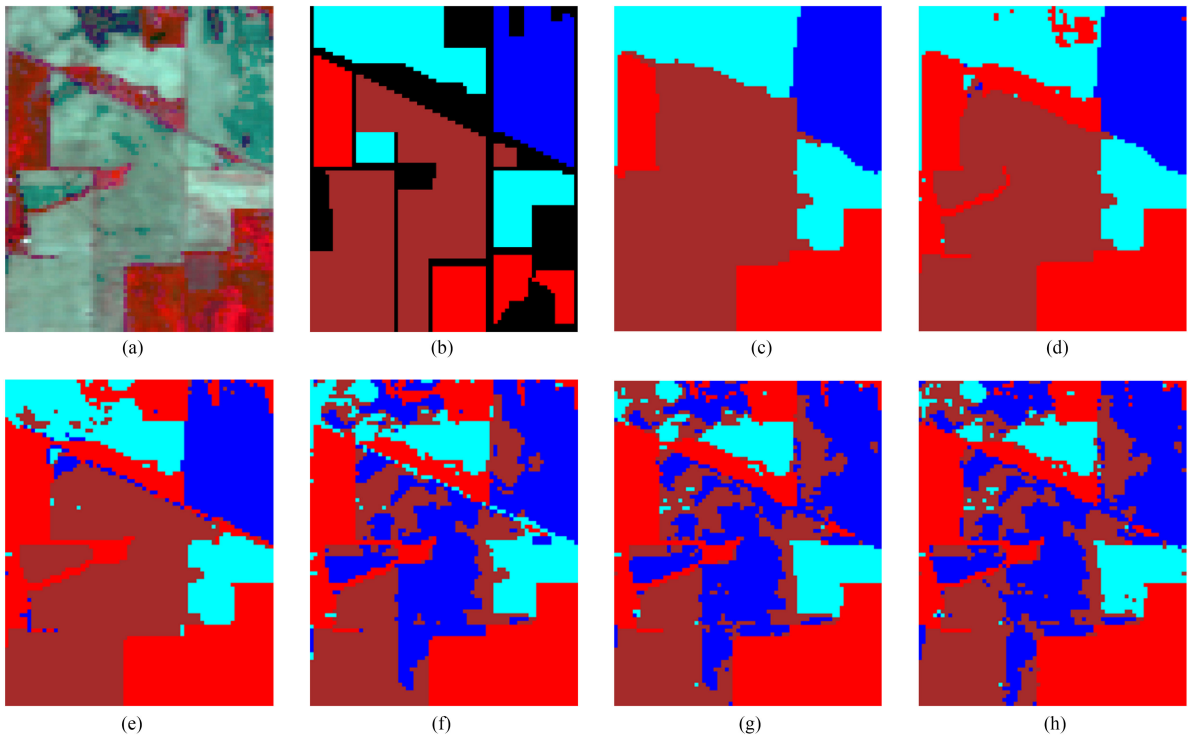

Fig. 6. *Indian Pines* image. (a) False color image, (b) ground truth, and clustering maps of JSSC-L obtained with (c) $p = 10$, (d) $p = 20$, (e) $p = 40$, (f) $p = 80$, (g) $p = 160$, and (h) $p = 320$.

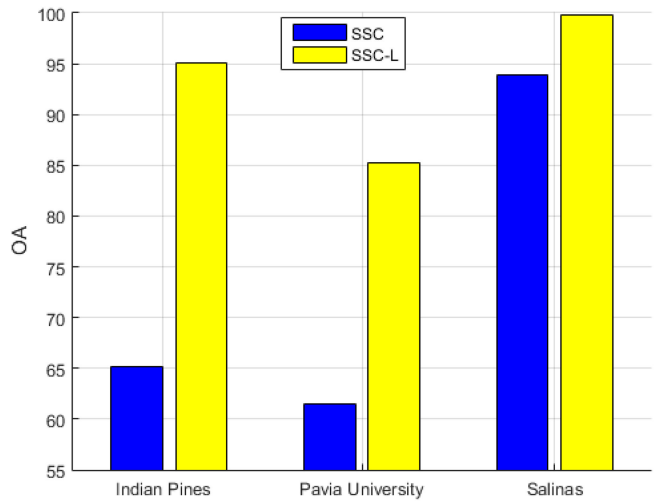

Fig. 7. Generalization of label information on SSC model in different data sets.

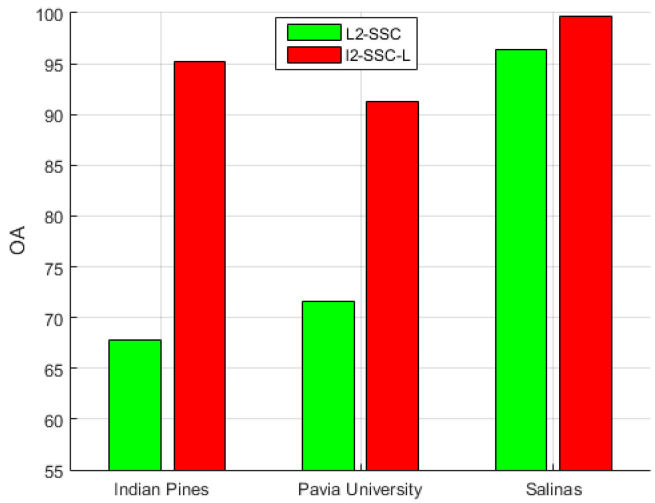

Fig. 8. Generalization of label information on L2-SSC model in different data sets.

we report the results in Fig. 9, where the standard deviations are denoted by the line width in different colors. In general, with increasing the fraction of labeled samples, the performance is consistently improved in all three data sets. The clustering accuracy shows first a drastic increase due to introducing labeled data, until about 1% of labeled samples per class are added, achieving 8.55%, 15.99%, and 2.01% OA improvements for the *Indian Pines*, *Pavia University*, and *Salinas*, respectively. Subsequently, the OA keeps increasing gently with further increase of the labeled data fraction. This demonstrates the effectiveness of our method in the case of limited number of labeled samples. For the *Salinas* image, the clustering accuracy reaches up to 100% with only 10% of labeled samples. The standard deviations at 1% labeled samples usually show larger values than that with more labeled samples, which is mainly caused by the spatial distribution of randomly selected samples.

G. Time Complexities: An Empirical Comparison

Table IV reports the computation times of different methods on the three HSIs. All the methods were implemented

TABLE IV
COMPARISONS BETWEEN DIFFERENT METHODS ON RUNNING TIME [IN SECONDS]

Data sets	FCM	k-means	CFSFDP	SSC	L2-SSC	CPPSSC	JSSC	JSSC-L
Indian Pines	6	**3**	9	543	624	342	270	304
Pavia University	19	**3**	80	17206	18048	11408	8380	8468
Salinas	10	**1**	18	1307	1459	784	610	650

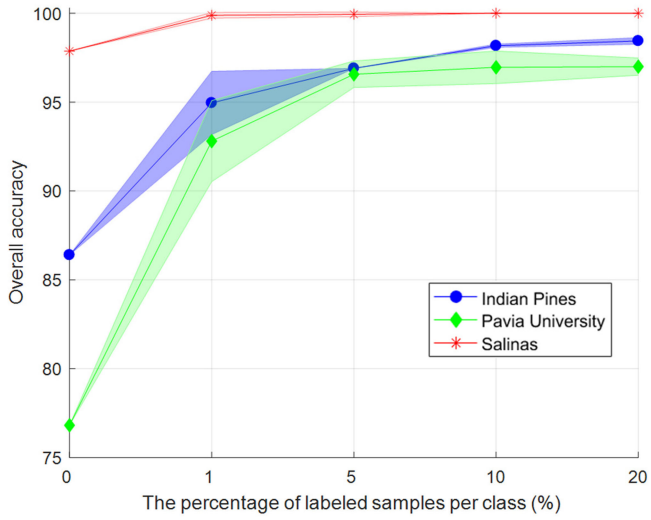

Fig. 9. Performance of JSSC-L on three different data sets, depending on the fraction of labeled samples.

in MATLAB and the experiments were run on a computer with an Intel core-i7 3930K CPU with 64 GB of RAM. The running time records the complete process of each clustering method. Clearly, the classical FCM and k-means methods and the CFSFDP method are much faster than all the SSC-based methods. However, the SSC-based methods typically achieve better results in terms of accuracy as shown above. CFSFDP takes longer time than FCM and k-means because of the preprocessing part regarding the calculation of pairwise distances for all the data points. Compared with SSC, L2-SSC, and CPPSSC, our methods JSSC and JSSC-L achieve faster convergence. The JSSC-L takes longer time than the JSSC, mainly due to the added constraint $\mathcal{P}_{\mathbb{L}}(\mathbf{C}) = \mathbf{0}$, but the increase of computation time is relatively small given the significantly enhanced clustering accuracy.

V. CONCLUSION

In this paper, we proposed a semisupervised SSC method for HSI, which combines spatial information via a joint sparsity constraint on local regions and label information within a unified framework. Our simple and elegant way of integrating the label information enables more accurate estimation of the similarity matrix and can be incorporated into other clustering methods as well, to generalize them to a semisupervised scenario. We also derived an optimization algorithm based on ADMM for our complete model. The experiments conducted on three real HSIs confirm the effectiveness of our method with a superior performance over other related clustering methods.

REFERENCES

[1] M. T. Eismann, A. D. Stocker, and N. M. Nasrabadi, "Automated hyperspectral cueing for civilian search and rescue," *Proc. IEEE*, vol. 97, no. 6, pp. 1031–1055, Jun. 2009.
[2] B. Du, Y. Zhang, L. Zhang, and D. Tao, "Beyond the sparsity-based target detector: A hybrid sparsity and statistics-based detector for hyperspectral images," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5345–5357, Nov. 2016.
[3] B. Datt, T. R. McVicar, T. G. Van Niel, D. L. Jupp, and J. S. Pearlman, "Preprocessing eo-1 hyperion hyperspectral data to support the application of agricultural indexes," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1246–1259, Jun. 2003.
[4] A. Bannari, A. Pacheco, K. Staenz, H. McNairn, and K. Omari, "Estimating and mapping crop residues cover on agricultural lands using hyperspectral and ikonos data," *Remote Sens. Environ.*, vol. 104, no. 4, pp. 447–459, 2006.
[5] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 45–54, Jan. 2014.
[6] C. Wu, L. Zhang, and B. Du, "Kernel slow feature analysis for scene change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 2367–2384, Apr. 2017.
[7] R. L. Lawrence, S. D. Wood, and R. L. Sheley, "Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest)," *Remote Sens. Environ.*, vol. 100, no. 3, pp. 356–362, 2006.
[8] H. Zhang, H. Zhai, L. Zhang, and P. Li, "Spectral–spatial sparse subspace clustering for hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3672–3684, Jun. 2016.
[9] S. Lloyd, "Least squares quantization in pcm," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 129–137, Mar. 1982.
[10] J. C. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*. New York, NY, USA: Springer, 1981.
[11] R. Vidal, Y. Ma, and S. Sastry, "Generalized principal component analysis (GPCA)," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 12, pp. 1945–1959, Dec. 2005.
[12] H. Zhang, Q. Wang, W. Shi, and M. Hao, "A novel adaptive fuzzy local information c-means clustering algorithm for remotely sensed imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5057–5068, Sep, 2017.
[13] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765–2781, Nov. 2013.
[14] W. Sun, L. Zhang, B. Du, W. Li, and Y. M. Lai, "Band selection using improved sparse subspace clustering for hyperspectral imagery classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2784–2797, Jun. 2015.
[15] H. Zhai, H. Zhang, L. Zhang, P. Li, and A. Plaza, "A new sparse subspace clustering algorithm for hyperspectral remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 1, pp. 43–47, Jan. 2017.
[16] H. Zhai, H. Zhang, X. Xu, L. Zhang, and P. Li, "Kernel sparse subspace clustering with a spatial max pooling operation for hyperspectral remote sensing data interpretation," *Remote Sens.*, vol. 9, no. 4, 2017, Art. no. 335.
[17] H. Zhai, H. Zhang, L. Zhang, and P. Li, "Total variation regularized collaborative representation clustering with a locally adaptive dictionary for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 166–180, Jan. 2019.
[18] H. Zhai, H. Zhang, L. Zhang, and P. Li, "Laplacian-regularized low-rank subspace clustering for hyperspectral image band selection," *IEEE Trans. Geosci. Remote Sens.*, to be published, 2018, doi: 10.1109/TGRS.2018.2868796.
[19] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Proc. Adv. Neural Inf. Process. Syst.*, 2002, pp. 849–856.

[20] H. Zhai, H. Zhang, L. Zhang, and P. Li, "Reweighted mass center based object-oriented sparse subspace clustering for hyperspectral images," *J. Appl. Remote Sens.*, vol. 10, no. 4, 2016, Art. no. 046014.
[21] Q. Yan, Y. Ding, Y. Xia, Y. Chong, and C. Zheng, "Class-probability propagation of supervised information based on sparse subspace clustering for hyperspectral images," *Remote Sens.*, vol. 9, no. 10, 2017, Art. no. 1017.
[22] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.
[23] H. Zhang, J. Li, Y. Huang, and L. Zhang, "A nonlocal weighted joint sparse representation classification method for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2056–2065, Jun. 2014.
[24] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral–spatial classification of hyperspectral images with a superpixel-based discriminative sparse model," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4186–4201, Aug. 2015.
[25] W. Fu, S. Li, L. Fang, X. Kang, and J. A. Benediktsson, "Hyperspectral image classification via shape-adaptive joint sparse representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 556–567, Feb. 2016.
[26] J. Liang, J. Zhou, Y. Qian, L. Wen, X. Bai, and Y. Gao, "On the sampling strategy for evaluation of spectral-spatial methods in hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 862–880, Feb. 2017.
[27] S. Huang, H. Zhang, and A. Pižurica, "A robust sparse representation model for hyperspectral image classification," *Sensors*, vol. 17, no. 9, 2017, Art. no. 2087.
[28] H. Zhai, H. Zhang, L. Zhang, and P. Li, "Cloud/shadow detection based on spectral indices for multi/hyperspectral optical remote sensing imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 144, pp. 235–253, 2018.
[29] F. Luo, B. Du, L. Zhang, L. Zhang, and D. Tao, "Feature learning using spatial-spectral hypergraph discriminant analysis for hyperspectral image," *IEEE Trans. Cybern.*, to be published, 2018, doi: 10.1109/TCYB.2018.2810806.
[30] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.
[31] E. Elhamifar and R. Vidal, "Robust classification using structured sparse representation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2011, pp. 1873–1879.
[32] S. Huang, H. Zhang, and A. Pižurica, "Joint sparsity based sparse subspace clustering for hyperspectral images," in *Proc. IEEE 25th Int. Conf. Image Process.*, 2018, pp. 3878–3882.
[33] E. Elhamifar and R. Vidal, "Sparse subspace clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2009, pp. 2790–2797.
[34] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.
[35] M. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2011, pp. 2097–2104.
[36] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.

Shaoguang Huang (S'17) received the B.S. degree in telecommunication engineering from Zhengzhou University, Zhengzhou, China, and the M.S. degree in telecommunication and information system from Shandong University, Jinan, China, in 2015. He is currently working toward the Ph.D. degree in computer science engineering from Ghent University, Ghent, Belgium.

His research interests include image processing, sparse representation, hyperspectral image analysis, and machine learning.

Hongyan Zhang (M'13–SM'16) received the B.S. degree in geographic information system and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2005 and 2010, respectively.

He has been a Full Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, since 2016. He is also a Young Chang-Jiang Scholar appointed by the Ministry of Education of China. He has authored/coauthored more than 80 research papers. His research interests include image reconstruction for quality improvement, hyperspectral information processing, and agricultural remote sensing.

Dr. Zhang serves as an Associate Editor of *Photogrammetric Engineering and Remote Sensing* and *Computers and Geosciences*. He is a Reviewer of more than 30 international academic journals, including IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, and IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.

Aleksandra Pižurica (SM'15) received the Diploma in electrical engineering from the University of Novi Sad, Novi Sad, Serbia, in 1994, the Master of Science degree in telecommunications from the University of Belgrade, Belgrade, Serbia, in 1997, and the Ph.D. degree in engineering from Ghent University, Ghent, Belgium, in 2002.

She is a Professor in statistical image modeling with Ghent University. Her research interests include the area of signal and image processing and machine learning, including multiresolution statistical image models, Markov Random Field models, sparse coding, representation learning, and image and video reconstruction, restoration, and analysis.

Prof. Pižurica serves currently as a Senior Area Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING and an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY. She was also an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING (2012–2016) and the Lead Guest Editor for the *EURASIP Journal on Advances in Signal Processing* for the Special Issue "Advanced Statistical Tools for Enhanced Quality Digital Imaging with Realistic Capture Models" (2013). The work of her team has been awarded twice the Best Paper Award of the IEEE Geoscience and Remote Sensing Society Data Fusion contest, in 2013 and 2014. She received the scientific prize "de Boelpaepe" for 2013–2014, awarded by the Royal Academy of Science, Letters and Fine Arts of Belgium for her contributions to statistical image modeling and applications to digital painting analysis.