

Locality and Structure Regularized Low Rank Representation for Hyperspectral Image Classification

Qi Wang[✉], Senior Member, IEEE, Xiang He, and Xuelong Li[✉], Fellow, IEEE

Abstract—Hyperspectral image (HSI) classification, which aims to assign an accurate label for hyperspectral pixels, has drawn great interest in recent years. Although low-rank representation (LRR) has been used to classify HSI, its ability to segment each class from the whole HSI data has not been exploited fully yet. LRR has a good capacity to capture the underlying low-dimensional subspaces embedded in original data. However, there are still two drawbacks for LRR. First, the LRR does not consider the local geometric structure within data, which makes the local correlation among neighboring data easily ignored. Second, the representation obtained by solving LRR is not discriminative enough to separate different data. In this paper, a novel locality- and structure-regularized LRR (LSLRR) model is proposed for HSI classification. To overcome the above-mentioned limitations, we present locality constraint criterion and structure preserving strategy to improve the classical LRR. Specifically, we introduce a new distance metric, which combines both spatial and spectral features, to explore the local similarity of pixels. Thus, the global and local structures of HSI data can be exploited sufficiently. In addition, we propose a structural constraint to make the representation have a near-block-diagonal structure. This helps to determine the final classification labels directly. Extensive experiments have been conducted on three popular HSI data sets. And the experimental results demonstrate that the proposed LSLRR outperforms other state-of-the-art methods.

Index Terms—Block-diagonal structure, hyperspectral image (HSI) classification, low-rank representation (LRR).

Manuscript received June 10, 2018; revised July 6, 2018; accepted July 22, 2018. Date of publication August 27, 2018; date of current version January 21, 2019. This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFB1002202, in part by the National Natural Science Foundation of China under Grant 61773316, in part by the Natural Science Foundation of Shaanxi Province under Grant 2018KJXX-024, in part by Fundamental Research Funds for the Central Universities under Grant 3102017AX010, and in part by the Open Research Fund of Key Laboratory of Spectral Imaging Technology Chinese Academy of Sciences. (Corresponding author: Qi Wang.)

Q. Wang is with the School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, China, also with the Center for Optical Imagery Analysis and Learning, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, China, and also with the Unmanned System Research Institute, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, China (e-mail: crabwq@gmail.com).

X. He is with the School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, China and also with Center for Optical Imagery Analysis and Learning, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, China (e-mail: xianghe@mail.nwpu.edu.cn).

X. Li is with the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China (e-mail: xuelongli@opt.ac.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/TGRS.2018.2862899

I. INTRODUCTION

HYPERSPECTRAL images (HSIs) are acquired by hyperspectral imaging sensors from the same spatial location and different spectral wavelengths. Due to the quite small wavelength interval (usually 10 nm) between every two neighboring bands, HSI generally has a very high spectral resolution. HSI is acquired from hundreds of continuous wavelengths, including a large range from visible to infrared spectrum; hence, HSI is composed of a great number of spectral bands, which makes hyperspectral data contain abundant discriminative information for the observed land surface. Since HSI can reflect well on the distinct property of different land materials, HSI classification [1], which is to assign the pixels of HSI a proper label, has attracted much attention over the past few decades.

Although the rich spectral information for each pixel brings a lot of help to classify hyperspectral data, there are still many challenges in HSI classification task. Due to the hundreds of spectral bands, the data of HSI has a very high dimensionality, which leads to the Hughes phenomenon [2]. In addition, it usually costs much time to label HSI data sets, hence most hyperspectral data have very limited training samples, which becomes another major challenge. To address the above-mentioned problems, a large number of support vector machine (SVM)-based approaches have been developed over the past years. SVM is a widely used classifier in most classification tasks. Since it can effectively handle the high-dimensional data, SVM has achieved great success in HSI classification. SVM with composite kernel (SVMCK) [3] was proposed to construct multiple composite kernels, which integrates both spectral and spatial information to enhance the classification performance. Specifically, the weighted kernels in [3] can effectively solve the problem that HSI usually has the limited labeled samples. In addition, SVM with graph kernel [4] developed a recursive graph kernel, which considered high-level spatial relationship rather than the simple pairwise relation. In addition to the advantage that graph kernel is easy to compute, it can also be suitable for the small training data. However, SVM-based methods have a common drawback that their performance is easy to be influenced by parameters settings.

Motivated by recent development in subspace segmentation, low-rank representation (LRR) has become an

effective method for HSI classification. LRR was first proposed for subspace segmentation by Liu *et al.* in [5]. Due to its considerable ability to exploit the underlying low-dimensional subspace structures of the given data, LRR has attracted extensive attention and achieved great success in various fields, such as face recognition [6], image classification [7], subspace clustering [8], object detection [9], and so on. In particular, LRR is also applied successfully in hyperspectral image (HSI) analysis [10] and obtains promising performance in the past few years. For instance, Sun *et al.* [11] presented a structured group low-rank prior, incorporating the spatial information, for sparse representation (SR) to classify HSI. Mei *et al.* [12] proposed to decompose the original hyperspectral data into the low-rank intrinsic spectral signature and sparse noise to alleviate spectral variation, which degrades strongly the performance of hyperspectral analysis. However, there are still some shortcomings for common LRR. First, in spite that LRR has a great ability to capture the global structure of the given data, it ignores the equally crucial local structure. This makes LRR fail to characterize the neighboring relation of each two pixels. Second, if all data are located in the union of multiple independent subspaces, the observed data with the same class should lie in the same subspace. Therefore, the ideal representation of the given data would have a class-wise block-diagonal structure. Nevertheless, the traditional LRR cannot obtain that structure. Third, most LRR-based methods employ the whole samples as the dictionary to learn the low-rank representation. However, the dictionary has too many redundant atoms, which not only increases the computational cost but also decreases the discriminative ability to reveal the potential property of HSI.

To tackle the aforementioned drawbacks, this paper proposes a novel locality- and structure-regularized LRR (LSLRR) for HSI classification. The main contributions are summarized as follows.

- 1) We introduce a new distance metric to measure the similarity of HSI pixels. For HSI classification, the spatial information is of great importance to acquire higher classification accuracy. The proposed measurement skillfully combines both spectral and spatial features into a unified distance metric, in which the involved parameter can be adjusted to fit different HSI data sets with different compactness of each class.
- 2) We present a novel locality constraint criterion (LCC) for LRR to further exploit the low-dimensional manifold structure of HSI. LRR can effectively capture the global structure of the given data, but the local geometry structure is also significant for most tasks. The proposed LSLRR with LCC successfully characterizes the global and local structures of HSI to explore the more reasonable representation.
- 3) An effective structure preserving strategy (SPS) is proposed to learn the more discriminative LRR for HSI data. As we all know, the ideal representation of multiclass data has a classwise block-diagonal structure. However, the original LRR hardly obtains the representation like that. Moreover, the learned representation for the testing set can be used directly to classify HSI.

The remainder of this paper is organized as follows. In Section II, two typical representation-based methods for HSI analysis are introduced. Then, the proposed LSLRR is described in detail in Section III. An optimization algorithm for solving LSLRR is derived in Section IV. In addition, Section V shows the extensive experimental results and corresponding analyses. Finally, we conclude this paper in Section VI.

II. RELATED WORK

As we all know, LRR and SR are two typical representation-based approaches. This paper mainly focuses on LRR, which has achieved huge success in hyperspectral remote sensing fields [13]. Since SR has some features common with LRR and has also attracted much attention in recent years, we will provide an overview of both SR-based and LRR-based methods for HSI classification in this section. In addition, the involved dictionary learning techniques are also introduced here.

A. SR-Based Methods

Given some data vectors, SR seeks the SR based on the linear combination of atoms in the dictionary. Due to its great classification performance, SR has been applied widely in the hyperspectral analysis. Chen *et al.* [14] proposed a joint sparsity model that represented the hyperspectral pixels within a patch by the same sparse coefficients. In [15], the sparsity of HSI was exploited by a probabilistic graphical model, which can effectively capture the conditional dependencies. Zhang *et al.* [16] developed a nonlocal weighted joint SR model, where different weights were employed to spatial neighboring pixels. In order to solve the problem that SR-based methods usually neglect the representation residuals, Li *et al.* [17] proposed a robust SR for HSI classification, which is robust for outliers. Moreover, Li *et al.* [18] presented a new superpixel-level joint sparse model for HSI classification, which explored the class-level sparsity to combine multiple features of pixels in local regions. A spectral-spatial adaptive SR was developed for HSI compression in [19], which made use of both spectral and spatial features, and it utilized superpixel segmentation to generate adaptive homogeneous regions. Gan *et al.* [20] incorporated multiple types of features into a kernel SR classifier, which is much helpful for HSI classification task. In addition, Fang *et al.* [21] proposed a multiscale adaptive SR that effectively integrated contextual feature at multiple scales by an adaptive sparse technique. Considering that ℓ_1 -based SR may obtain unstable representation results, Tang *et al.* [22] incorporated manifold learning into SR to exploit the local structure and get the smooth sample representation. For a more detailed description, a useful survey about SR-based methods can be referred in [23].

B. LRR-Based Methods

Another popular representation-based method is LRR. Different from SR, LRR seeks low-rank representation for the given data, and most LRR-based approaches have been proposed for HSI analysis [24]. Du *et al.* [25] utilized the joint sparse and LRR to solve the abundance estimation problem for HSI. The low-rank constraint is integrated to overcome the drawback of local spectral redundancy and

correlation for HSI denoising in [26]. Shi *et al.* [27] proposed a semisupervised framework for HSI classification, where LRR reconstruction is employed to decrease the influence of noise and outliers and make domain adaption more robust. A novel framework combining the maximum a posteriori and LRR, exploiting the high spectral correlation, is proposed for HSI segmentation in [28]. Considering that the underlying low-dimensional structure in HSI data is multiple subspaces rather than single subspace, Sumarsono *et al.* [29] adopted LRR as a preprocessing step for supervised and unsupervised classification of HSI. Most studies have demonstrated that the contextual information is very beneficial to improve the classification accuracy of HSI. Almost all the state-of-the-art work, which employed LRR for HSI classification, combined both spectral and spatial features. For instance, a new low-rank structured group prior was presented to exploit the spatial information between neighboring pixels by Sun *et al.* [11]. Soltani-Farani *et al.* [30] proposed to add the spatial characteristics by partitioning the HSI into several square patches as contextual groups. However, the fixed-size squares window neglects the difference between the pixels in the same window. He *et al.* [31] applied a superpixel segmentation algorithm to divide HSI into some homogeneous regions with adaptive size, which is better than fixed-size patches to utilize the contextual features. In addition, a new spectral-spatial HSI classification method using $\ell_{1/2}$ -regularized LRR was developed in [32], where the contextual information is efficiently incorporated into the spectral signatures by representing the spatial adjacent pixels in a low-rank form.

C. Dictionary Learning

Since LRR can greatly exploit the global structure for the given data, it is superior to SR in some cases. However, one thing that LRR and SR have in common is that they both assume to describe every sample as the linear combination of some atoms in a given dictionary, and the selection of dictionary is fairly important to the performance of LRR. In general, dictionary learning methods can be roughly divided into two categories [33]: 1) learning a dictionary based on the mathematical model. Many traditional models such as contourlet, wavelet, bandelet, wavelet packets, etc., all can be used to construct an effective dictionary. 2) building a dictionary to behave well in training set. The second class of methods has brought more and more concern. The major advantage is that they can obtain great experimental results in most practical applications. These state-of-the-art methods include Optimal Directions [34], Union of Orthobases [35], Generalized principal component analysis [36], K-singular value decomposition (SVD) [37] and so on. For HSI classification, some dictionary learning techniques have been proposed. Soltani-Farani *et al.* [30] presented a spatial-aware dictionary learning method that is to divide HSI data into some contextual neighborhoods and then model the pixels with the same group as a common subspace. Motivated by Learning Vector Quantization, Wang *et al.* [38] proposed a novel dictionary learning method for the SR, and modeled the spatial context

by a Bayesian graph. He *et al.* [31] applied a joint LRR model in every spatial group to learn an appropriate dictionary.

III. LOCALITY- AND STRUCTURE-REGULARIZED LOW-RANK REPRESENTATION

In this section, we will describe the proposed LSLRR in detail. The original LRR formulas are first introduced. Then, two main powerful regularization terms and dictionary learning scheme are presented. Finally, we derive an optimization algorithm to solve the objective function of LSLRR.

A. Low-Rank Representation

LRR is based on the assumption that all data are sufficiently sampled from multiple low-dimensional subspaces embedded in a high-dimensional space. Reference [5] indicates that LRR can effectively explore the underlying low-dimensional structures for the given data. Assume that data samples $Y \in \mathbb{R}^{d \times n}$ are drawn from a union of many subspaces that are denoted as $\bigcup_{i=1}^k S_k$, where S_1, S_2, \dots, S_k are the low-dimensional subspaces. The LRR model aims to seek the low-rank representation $Z \in \mathbb{R}^{m \times n}$ and the sparse noises $E \in \mathbb{R}^{d \times n}$ based on the given dictionary $A \in \mathbb{R}^{d \times m}$. Specifically, LRR is formulated as the following rank minimization problem:

$$\min_{Z, E} \text{rank}(Z) + \lambda \|E\|_0 \quad \text{s.t. } Y = AZ + E \quad (1)$$

where A and E are the dictionary matrix and sparse noise component, respectively. $\|\cdot\|_0$ is the ℓ_0 -norm, the number of all nonzero elements. λ is the regularization coefficient to balance the weights of rank term and reconstruction error. It is worth noting that the only difference between SR and LRR is that SR aims to find the sparsest representation while LRR seeks the low-rank representation, and can effectively capture the global structure of data samples.

However, it is difficult to solve the nonconvex problem (1) due to the discrete nature of the rank operation and ℓ_0 -norm. Therefore, the original minimization problem (1) needs to be relaxed in order to make it solvable. The common convex relaxation of problem (1) is presented as

$$\min_{Z, E} \|Z\|_* + \lambda \|E\|_1 \quad \text{s.t. } Y = AZ + E, \quad (2)$$

where $\|\cdot\|_*$, defined as the sum of all singular values of Z , is the nuclear norm; $\|\cdot\|_1$ is the ℓ_1 -norm, i.e., the sum of the absolute value of all elements; and $\|Z\|_*$ and $\|E\|_1$ are the convex envelope of $\text{rank}(Z)$ and $\|E\|_0$, respectively. Then, problem (2) has a nontrivial solution. In fact, the solution of problem (2) is equal to that of problem (1) in this case of free noise [8]. However, in practical applications, most data are noisy, even strongly corrupted. Therefore, when a large number of data samples are grossly corrupted, a robust model [5] is presented as

$$\min_{Z, E} \|Z\|_* + \lambda \|E\|_{2,1} \quad \text{s.t. } Y = AZ + E \quad (3)$$

where $\|\cdot\|_{2,1}$ is the $\ell_{2,1}$ -norm, which is defined as $\|E\|_{2,1} = \sum_j^n (\sum_i^d E_{i,j}^2)^{1/2}$. Specifically, compared to ℓ_1 -norm, $\ell_{2,1}$ -norm expects more columns of E to be zero vector, i.e., some samples are clean and others are noisy.

B. Locality Constraint Criterion for LSLRR

For HSI classification, if some pixels have a neighboring relation, there is a high probability that they belong to the same class. That is, the spatial similarity is a beneficial information to improve the classification accuracy of HSI. Therefore, it is very necessary to incorporate the contextual information into the classifier. Furthermore, LRR has a powerful ability to exploit the global structure of HSI data, but the local manifold structure between the adjacent pixels, which is also helpful to classify HSI, is neglected by LRR. Therefore, we develop a local structure constraint, which utilizes both the spectral and spatial similarity, to improve the performance of the original LRR model.

Suppose that HSI data are denoted as $X = [x_1, x_2, \dots, x_n] \in \mathbb{R}^{d \times n}$, where d and n are the number of spectral bands and all pixels, respectively, and x_i denotes the spectral column vector of the i th pixel of HSI data X . Similarly, assume that the spatial feature matrix $L = [l_1, l_2, \dots, l_n] \in \mathbb{R}^{2 \times n}$, and l_i denotes the position coordinate of the i th pixel. A simple way to compute the distance matrix which combines both spectral and spatial features is formulated as

$$M_{ij} = \sqrt{\|x_i - x_j\|_2^2 + \|l_i - l_j\|_2^2}, \quad (4)$$

where M_{ij} is the distance between the i th and j th pixels. Note that the spectral values of X and coordinate values of L are normalized to a range of $[0, 1]$. However, the distance metric is not reasonable enough because the above-mentioned spectral and spatial features are unequal and have different physical meanings. Therefore, a more accurate similarity metric between two pixels is proposed as

$$M_{ij} = \sqrt{\|x_i - x_j\|_2^2 + m\|l_i - l_j\|_2^2} \quad (5)$$

where m is a hyperparameter for controlling the weight of spectral and spatial distance. For different HSI data sets, the compactness of each category is different, and it is more appropriate to choose a large value of m for the HSI data set with high compactness of each class. As we all know, two pixels with a larger distance should have a smaller similarity. In addition, the low-rank representation Z can be viewed as the affinity matrix, in which Z_{ij} denotes the similarity of the i th and j th samples. As such, to keep the difference between classes and the compactness within classes, the locality constraint as a penalty term for LRR is introduced as follows:

$$\sum_{i,j} M_{ij} |Z_{ij}| = \|M \circ Z\|_1 \quad (6)$$

where \circ is the Hadamard product that denotes the elementwise product of two matrices. Moreover, the locality constraint also takes the sparsity of low-rank representation matrix Z into account. Because Z stands for the similarity between dictionary and the original data, all elements of Z should have nonnegative values. Therefore, the final locality regularization term can be written as $\|M \circ Z\|_1$ with the constraint $Z \geq 0$, and locality-regularized LRR model can be formulated as

$$\begin{aligned} \min_{Z, E} \quad & \|Z\|_* + \lambda \|E\|_{2,1} + \alpha \|M \circ Z\|_1 \\ \text{s.t.} \quad & Y = AZ + E, \quad Z \geq 0. \end{aligned} \quad (7)$$

C. Structure Preserving Strategy for LSLRR

Hyperspectral data X are first divided into two parts, denoting $X = [\bar{X}, \hat{X}]$, where \bar{X} represents the training data and \hat{X} represents the testing data. Rearrange the permutation of samples according to each class that $\bar{X} = [\bar{X}_1, \bar{X}_2, \dots, \bar{X}_c] \in \mathbb{R}^{d \times m}$, where X_i is the i th class set of training samples, and c denotes the number of classes. In addition, $\hat{X} = [\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n] \in \mathbb{R}^{d \times n}$ is the testing feature matrix, whose i th column is the spectral vector of the i th testing sample. In LRR model, we set the data $Y = [\bar{X}, \hat{X}]$ while the dictionary $A = \bar{X}$. Therefore, $[\bar{X}, \hat{X}] = \bar{X}Z$ is obtained. Similarly, Z can be written as $[\bar{Z}, \hat{Z}]$, where \bar{Z} and \hat{Z} are the low-rank representations for \bar{X} and \hat{X} under the base \bar{X} , respectively.

In general LRR model, all data are used as the dictionary and each sample is considered as the atom of the dictionary, e.g., $X = XZ + E$. When removing sparse noise E , the data X can be reconstructed by low-rank representation Z based on the data itself. Furthermore, if data samples are permuted based on the order of classes, the ideal representation matrix Z would have a classwise block-diagonal structure as follows:

$$Z = \begin{bmatrix} Z_1^* & 0 & 0 & 0 \\ 0 & Z_2^* & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & Z_c^* \end{bmatrix} \quad (8)$$

where c is the number of classes. The proposed model $[\bar{X}, \hat{X}] = \bar{X}[\bar{Z}, \hat{Z}] + E$ has a similar property to the classical LRR model $X = XZ + E$. That is, the representation matrices \bar{Z} and \hat{Z} should also have a classwise block-diagonal structure similar to the form of (8).

To make \bar{Z} and \hat{Z} hold the above-mentioned structure, we introduce a structured auxiliary matrix \hat{Q} to constrain Z . First, \hat{Q} is also divided into two parts: \bar{Q} and \hat{Q} . We can obtain $\bar{Z}_i^*, i = 1, 2, \dots, c$, with setting $A = X_i$ by solving the model (7). Let $\bar{Q} = \text{diag}(\bar{Z}_1^*, \bar{Z}_2^*, \dots, \bar{Z}_n^*)$, where diag is the diagonal operation. Note that this step actually utilizes the label information for the training data \bar{X} . Therefore, the classwise block-diagonal structure for \bar{Z} is easy to preserve. Second, it is difficult to hold the structure (8) for \hat{Z} without a prior about the number of each class testing samples. As we already know, there are many zero elements in \hat{Z} when it has a block-diagonal structure. In addition, we previously mentioned that Z_{ij} represents the similarity of the i th and j th samples. We employ the Gaussian similarity function to generate the auxiliary matrix \hat{Q} as follows:

$$\hat{Q}_{ij} = \exp\left(-\frac{\|x_i - x_j\|_2^2 + m\|l_i - l_j\|_2^2}{\sigma}\right) \quad (9)$$

where the parameter σ is used to control the width of neighbors. If the distance between the i th training pixel and the j th testing pixel is large enough (e.g., larger than θ , where θ is the maximum distance parameter), we will set $\|x_i - x_j\|_2^2 + m\|l_i - l_j\|_2^2 = \infty$. Thus, \hat{Q} would have many zero elements and \hat{Z} would be a sparse matrix. Finally, \hat{Q} is obtained by $\hat{Q} = [\bar{Q}, \hat{Q}]$. Therefore, the structure constraint can be written as $\|Z - \hat{Q}\|_F^2$, which makes the low-rank

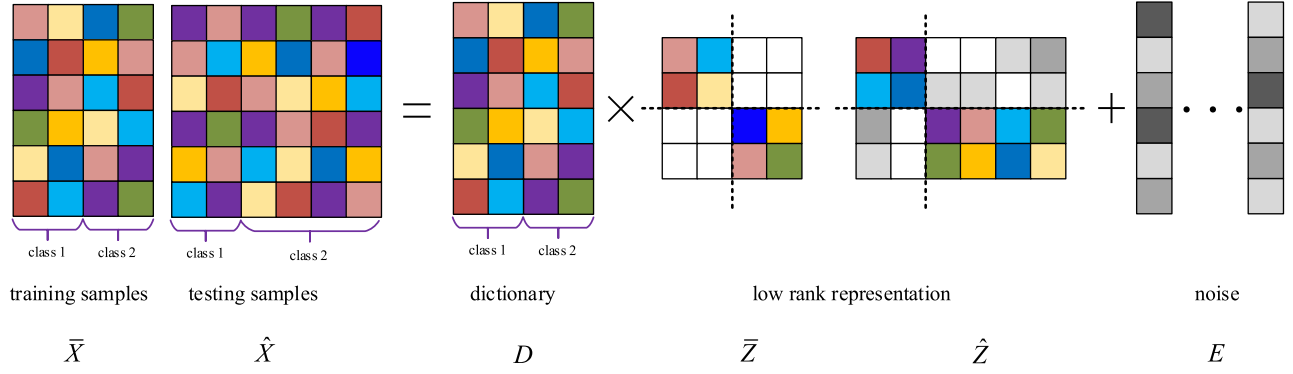


Fig. 1. Illustration of the proposed LSLRR method. The value of colored and white blocks are nonzero and zero, respectively. In addition, the value of gray blocks is close to zero. For the purpose of simplification, only two classes are used to describe the method.

representation \bar{Z} and \hat{Z} have an approximately block-diagonal structure.

Considering that the j th column of Z represents the similarity between each training pixels and the j th testing pixel, we enforce the sum of each column of Z to be 1, i.e., $1_m^T Z = 1_{m+n}^T$. After incorporating the aforementioned two crucial techniques into the classical LRR model, the LSLRR can be formulated as

$$\begin{aligned} \min_{Z, E} \quad & \|Z\|_* + \lambda \|E\|_{2,1} + \alpha \|M \circ Z\|_1 + \beta \|Z - Q\|_F^2 \\ \text{s.t.} \quad & X = \bar{X}Z + E, \quad 1_m^T Z = 1_{m+n}^T, \quad Z \geq 0 \end{aligned} \quad (10)$$

where 1_m and 1_{m+n} are the unit vectors with length of m and $m+n$, respectively.

D. Dictionary Learning for LSLRR

Dictionary learning is a crucial step for most classification problems. Generally, the whole samples are usually used for the dictionary for LRR. However, when the data samples are corrupted by noise, they cannot reconstruct themselves well by the polluted dictionary. In addition, the high-quality dictionary can improve significantly the performance of the classification methods. The process of learning the LRR can also become easy with a compact dictionary. Here, we will learn a discriminative dictionary from the corrupted HSI data.

For the problem (10), the dictionary is randomly selected from HSI data, and the atoms in \bar{X} are a part of the whole HSI pixels. In the solving process, the dictionary \bar{X} is fixed. However, if the selected samples are not representative and discriminative, or even worse (i.e., grossly corrupted) for the whole data, the obtained low-rank representation Z would be useless. Therefore, we integrate a dictionary learning process into the problem (10) instead of fixing some dictionary atoms. Then, the final objective function can be demonstrated as

$$\begin{aligned} \min_{Z, E, D} \quad & \|Z\|_* + \lambda \|E\|_{2,1} + \alpha \|M \circ Z\|_1 + \beta \|Z - Q\|_F^2 \\ \text{s.t.} \quad & X = DZ + E, \quad 1_m^T Z = 1_{m+n}^T, \quad Z \geq 0 \end{aligned} \quad (11)$$

where α and β control the weights of locality and structure constraints, respectively. The proposed method, namely

LSLRR, has a considerable ability to require the block-diagonal representation and simultaneously to learn a discriminative dictionary. In addition, Fig. 1 illustrates the proposed LSLRR. The given data are first divided into training set \bar{X} and testing set \hat{X} . Then, the LRR matrix \bar{Z} for the training set and \hat{Z} for the testing set are obtained based on the dictionary D . In addition, \bar{Z} is a block-diagonal matrix, and \hat{Z} is an approximate block-diagonal matrix.

E. HSI Classification Via LSLRR

Hyperspectral pixels belonging to the same class have an extremely similar spectral reflectance curve, which is the theoretical evidence to classify HSI. Although HSI data have a large number of bands and the dimensionality is very high, the similarity between neighboring bands is also very high. Reference [39] indicates that many low-dimensional subspaces exist in HSI data space. In addition, Chakrabarti *et al.* [40] made a lot of statistical analyses based on the real-world HSI data, and came to a conclusion that the rank of HSI data matrix is approximately equal to the number of classes. This implies that HSI data satisfy the low-rank property. Pixels of each class have a similar position in the whole HSI space, and they make up a low-dimensional subspace. For the proposed LSLRR, it can effectively segment these subspaces embedded in HSI from both global and local aspects. Recall that \hat{z}_{ij} in \hat{Z} stands for the similarity of the i th training pixel and j th testing pixel. The larger the value of \hat{z}_{ij} is, the higher the possibility of x_i and x_j belongs to the same class. Therefore, the final classification results can be directly obtained and it is no need to employ some complex classification algorithms. Specifically, the label of a testing pixel x_j can be confirmed as follows. First, compute the sum of the j th column of \hat{Z} for each class. The result is denoted by $S_l(\hat{z}_j)$, $l \in [1, \dots, c]$. Second, the label of x_j , denoted by $\text{label}(x_j)$, is determined as

$$\text{label}(x_j) = \arg \max_{l=1, \dots, c} S_l(\hat{z}_j). \quad (12)$$

IV. OPTIMIZATION ALGORITHM FOR SOLVING LSLRR

In this section, we derive an optimization algorithm to solve the LSLRR model (11). In recent years, a large number of

algorithms [41], [42] have been developed to solve the rank minimization optimization problem. Here, we adopt the high-efficiency inexact augmented Lagrange Multiplier (IALM) method to solve the proposed LSLRR. First, we introduce two auxiliary variables H and J to make the problem (11) become easily solvable. Thus, the equivalent problem of (11) is converted to

$$\begin{aligned} \min_{H, J, Z, E, D} \quad & \|Z\|_* + \lambda \|E\|_{2,1} + \alpha \|M \circ Z\|_1 + \beta \|Z - Q\|_F^2 \\ \text{s.t.} \quad & X = DZ + E, \quad Z = J, \quad H = Z, \\ & 1_m^T Z = 1_{m+n}^T, \quad Z \geq 0. \end{aligned} \quad (13)$$

Then, the corresponding augmented Lagrangian function for (13) can be written as

$$\begin{aligned} \min_{H \geq 0, J, Z, E, D} \quad & \|J\|_* + \lambda \|E\|_{2,1} + \alpha \|M \circ H\|_1 + \beta \|Z - Q\|_F^2 \\ & + \langle Y_1, X - DZ - E \rangle + \langle Y_2, Z - J \rangle \\ & + \langle Y_3, H - Z \rangle + \langle Y_4, 1_m^T Z - 1_{m+n}^T \rangle \\ & + \frac{\mu}{2} (\|X - DZ - E\|_F^2 + \|Z - J\|_F^2 \\ & + \|H - Z\|_F^2 + \|1_m^T Z - 1_{m+n}^T\|_F^2) \end{aligned} \quad (14)$$

where $\langle A, B \rangle = \text{trace}(A^T B)$, $\mu > 0$ is a penalty parameter and Y_1, Y_2, Y_3 , and Y_4 are the Lagrange multipliers. The alternative optimization algorithm can be applied to solve the problem (14) with five optimization variables (H, J, Z, E, D). The detailed updating schemes can be seen as follows.

Update H: fix J, Z, E , and D , and then H can be updated as follows:

$$H^{k+1} = \arg \min_{H \geq 0} \frac{\alpha}{\mu^k} \|M \circ H^k\|_1 + \frac{1}{2} \left\| H^k - Z^k + \frac{Y_3^k}{\mu^k} \right\|_F^2. \quad (15)$$

The solution for (15) can be computed [43] by

$$H_{ij}^{k+1} = \max \left[0, \Theta_{w_{ij}} \left(Z_{ij}^k - \frac{Y_{3,ij}^k}{\mu^k} \right) \right] \quad (16)$$

where $\Theta_w(x) = \max(x - w, 0) + \min(x + w, 0)$, $w_{ij} = (\alpha/\mu^k)M_{ij}$.

Update J: fix H, Z, E , and D , and then J can be updated as follows:

$$\begin{aligned} J^{k+1} &= \arg \min_J \frac{1}{\mu^k} \|J^k\|_* + \frac{1}{2} \|Z^k - J^k + \frac{Y_2^k}{\mu^k}\|_F^2 \\ &= US_{1/\mu^k}(\Sigma)V^T, \end{aligned} \quad (17)$$

where $U\Sigma V^T$ is the SVD of $Z^k + Y_2^k/\mu^k$, and $S_\epsilon(x) = \text{sgn}(x)\max(|x| - \epsilon, 0)$ is the soft-thresholding operator [5].

Update Z: fix H, J, E , and D , and then Z can be updated as follows:

$$\begin{aligned} Z^{k+1} &= \arg \min_Z \beta \|Z^k - Q\|_F^2 + \frac{\mu^k}{2} \left\| Z^k - J^k + \frac{Y_2^k}{\mu^k} \right\|_F^2 \\ &+ \frac{\mu^k}{2} \left\| X - DZ^k - E^k + \frac{Y_1^k}{\mu^k} \right\|_F^2 \\ &+ \frac{\mu^k}{2} \left\| H^k - Z^k + \frac{Y_3^k}{\mu^k} \right\|_F^2 \\ &+ \frac{\mu^k}{2} \left\| 1_m^T Z^k - 1_{m+n}^T + \frac{Y_4^k}{\mu^k} \right\|_F^2. \end{aligned} \quad (18)$$

Problem (18) is a quadratic minimization problem. And it has a closed-form solution, which can be obtained by making the derivative of (18) be zero. The optimal solution for variable Z is

$$Z^{k+1} = [W^k]^{-1} [2\beta Q^k + \mu^k (D^T A^k + B^k + C^k + 1_m F^k)], \quad (19)$$

where $A = X - E + Y_1/\mu$, $B = J - Y_2/\mu$, $C = H + Y_3/\mu$, $F = 1_{m+n}^T - Y_4/\mu$, and $W = 2\beta I + \mu(D^T D + 2I + 1_m 1_m^T)$.

Update E: fix H, J, Z , and D , and then E can be updated as follows:

$$E^{k+1} = \arg \min_E \frac{\lambda}{\mu^k} \|E^k\|_{2,1} + \frac{1}{2} \|X - DZ^k - E^k + \frac{Y_1^k}{\mu^k}\|_F^2. \quad (20)$$

Denote $G = X - DZ + Y_1/\mu$, then the j th column of optimal E [5] is

$$E^{k+1}(:, i) = \begin{cases} \frac{\|g_i\|_2 - \frac{\lambda}{\mu^k}}{\|g_i\|_2} g_i, & \text{if } \frac{\lambda}{\mu^k} < \|g_i\|_2 \\ 0, & \text{otherwise} \end{cases} \quad (21)$$

Update D: fix H, J, Z , and E , and then D can be updated as follows:

$$D^{k+1} = \arg \min_D \frac{\mu^k}{2} \left\| X - D^k Z^k - E^k + \frac{Y_1^k}{\mu^k} \right\|_F^2. \quad (22)$$

Problem (22) is also a quadratic minimization problem. Here, we employ an iteration updating strategy to obtain the optimal solution of dictionary D . First, we initialize the dictionary D^0 by randomly selecting a part of HSI pixels. Second, the updating dictionary D^{new} is obtained by solving the problem (22). Finally, the detailed updating rule is

$$D^{k+1} = w D^k + (1 - w) D^{\text{new}} \quad (23)$$

where w is a weight parameter. For each iteration, $D^{\text{new}} = (X - E + Y_1^k/\mu^k)Z^T(ZZ^T)^{-1}$.

Finally, the overall optimization algorithm for solving the proposed LSLRR (11) is described as Algorithm 1.

V. EXPERIMENTS AND ANALYSES

In this section, some comprehensive experiments are conducted to prove the effectiveness of the proposed LSLRR for HSI classification. Many state-of-the-art classification algorithms are considered as the comparison methods. After the experiments, some detailed analyses are also given.

Algorithm 1 IALM for Solving LSLRR

Input: testing set \hat{X} , training set \bar{X} , local constraint matrix M , structure constraint matrix Q , parameter $\lambda, \alpha, \beta, m$.

Output: low rank representation Z , the noise E .

Initialize: $H = J = Z = E = 0, D^0 = \bar{X}, \mu = 10^{-6}, \max_{\mu} = 10^{10}, \rho = 1.1, \varepsilon = 10^{-4}, Y_1 = Y_2 = Y_3 = Y_4 = 0$.

While not converged **do**

- 1) Compute the optimal solution of H, J, Z, E and D according to (16), (17), (19), (21), (23), respectively.
- 2) Update the Lagrange multipliers by

$$Y_1^{k+1} = Y_1^k + \mu^k (X - \bar{X}Z^k - E^k),$$

$$Y_2^{k+1} = Y_2^k + \mu^k (Z^k - J^k),$$

$$Y_3^{k+1} = Y_3^k + \mu^k (H^k - Z^k),$$

$$Y_4^{k+1} = Y_4^k + \mu^k (1_m^T Z^k - 1_{m+n}^T).$$
- 3) Update the parameter μ by $\mu^{k+1} = \max(\rho\mu^k, \max_{\mu})$.
- 4) Check the convergence conditions

$$\|X - DZ - E\|_{\infty} < \varepsilon, \|Z - J\|_{\infty} < \varepsilon,$$

$$\|H - Z\|_{\infty} < \varepsilon, \|D^{k+1} - D^k\|_{\infty} < \varepsilon,$$

$$\|1_m^T Z - 1_{m+n}^T\|_{\infty} < \varepsilon.$$
- 5) $k \leftarrow k + 1$.

End while

A. Data Set Descriptions

To evaluate the classification performance of the proposed LSLRR model, three popular hyperspectral data sets are used to conduct the verification experiments. The detailed descriptions are shown as follows [44].

- 1) *Indian Pines*: The scene is collected by airborne visible infrared imaging spectrometer (AVIRIS) sensor over the most agricultural regions in northwestern Indiana, USA. And the data set is composed of 145×145 pixels with 220 spectral bands whose wavelength ranges from 0.4–2.5 μm . After removing some noise and water-absorption bands, the remaining image has 200 spectral bands, which can be used for the classification task. In addition, there are 16 classes for this data set.
- 2) *Pavia University*: This data set was captured by reflective optics system imaging spectrometer sensor over the urban area of the University of Pavia, northern Italy, on July 8, 2002. The original data set consists of 115 spectral bands covering 0.43–0.86 μm , of which 12 noisy bands are removed and 103 bands are retained. The size of each band is 610×340 with a spatial resolution of 1.3 m per pixel. Nine categories of ground covering are considered for the classification experiments.
- 3) *Salinas*: The image is also gathered by AVIRIS sensor and contains the wavelength range of 0.4–2.5 μm like the Indian Pines. It has a high spatial resolution of 3.7 m per pixel. The covered area consists of 512 lines and 217 samples. In addition, there are 204 spectral bands after discarding some polluted bands. The number of

the ground category is also 16. This scene mainly consists of bare soils, vegetables, and vineyard fields.

B. Experimental Setups

Before demonstrating the experimental results, the comparison methods, corresponding parameter settings, and the evaluation indexes are first introduced as follows.

1) *Comparison Algorithms*: To verify the superiority of the proposed LSLRR, some state-of-the-art HSI classification methods are considered. They are: 1) SVM [45]; 2) SVMCK [3]; 3) joint robust SR classifier (JRSRC) [17]; 4) class-dependent SR classifier(cdSRC) [46]; 5) LRR [5]; and 6) low-rank group inspired dictionary learning (LGIDL). [31].

The above-mentioned competitors can roughly be divided into three categories: SVM-based, SR-based, and LRR-based methods. To be specific, the classic SVM is a great classifier that has been widely applied in HSI classification. And another powerful SVM-based method, SVMCK, has achieved promising classification accuracy due to incorporating the contextual information into the kernels. Furthermore, we also take two SR-based classification algorithms into account. The first one is JRSRC, which makes these pixels in neighboring regions represented jointly by some common training samples with the same sparse coefficients. An advantage for JRSRC is that it is robust to the HSI outliers. The second is cdSRC, which effectively integrates the idea of k-nearest neighbor into SRC in a classwise manner and characterizes both Euclidean distance and correlation information between training and testing set. Finally, these LRR-based approaches are the original LRR and LGIDL. Among them, the LGIDL employs superpixel segmentation to obtain the adaptive spatial correlation regions and yields fairly competitive performance.

2) *Parameter Settings*: Every method is repeated 10 times to avoid the bias due to the random sampling. All free parameters of these algorithms are determined via cross validation, using training data only. For SVM-based comparison methods, we choose radial basis function $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$ as the kernel function of SVM, and the optimal parameters C and γ are tuned by grid search algorithm. The one-versus-one strategy is applied in the implementation of SVM. Specifically, the parameters of SVM are $C = 2000, \gamma = 0.1$ for Indian Pines, $C = 1500, \gamma = 0.08$ for Pavia University, and $C = 4000, \gamma = 0.001$ for Salinas. For SVMCK, we select the mean spectral values of square patches as the spatial feature, and employ the weighted summation kernel to balance the spatial and spectral components. The patch size T and kernel weight μ for three data sets are $\{T = 15, \mu = 0.7\}, \{T = 5, \mu = 0.8\}$, and $\{T = 50, \mu = 0.4\}$. Moreover, the optimal parameter settings of JRSRC and LGIDL are followed as [17] and [31], respectively. For the proposed LSLRR, the corresponding parameters are set as $\{\lambda = 20, \alpha = 0.8, \beta = 0.6, m = 25\}, \{\lambda = 10, \alpha = 0.3, \beta = 1.2, m = 15\}, \{\lambda = 10, \alpha = 1, \beta = 0.4, m = 40\}$ for three HSI data sets, respectively.

3) *Evaluation Indexes*: We adopt three quantitative metrics, namely, overall accuracy (OA), average accuracy (AA), and

TABLE I
CLASSIFICATION ACCURACY (%) OF DIFFERENT COMPARISON METHODS AND THE PROPOSED LSLRR FOR INDIAN PINES DATA SET

Class	SVM	SVMCK	JRSRC	cdSRC	LRR	LGIDL	LSLRR
1	87.80	70.73	58.54	85.37	19.15	63.41	100
2	78.29	88.17	92.68	91.05	63.04	96.26	95.69
3	64.66	88.35	95.18	91.16	56.36	91.97	94.25
4	77.46	91.08	92.96	94.84	21.13	87.32	98.05
5	91.72	88.74	88.51	92.18	75.17	90.80	94.42
6	97.41	97.72	87.21	99.39	86.15	99.09	98.93
7	64.02	100	72.00	100	47.97	84.01	100
8	98.14	98.14	99.07	100	79.53	96.98	100
9	33.33	38.89	33.33	50.00	11.11	38.89	27.78
10	70.15	88.23	84.91	89.83	72.11	90.17	90.37
11	83.52	96.06	97.56	95.97	83.88	95.74	95.79
12	66.88	81.84	82.02	85.39	37.45	90.26	95.51
13	95.65	90.76	88.04	94.57	80.43	94.02	96.20
14	94.82	98.86	95.69	98.95	90.86	99.21	99.56
15	59.37	81.27	95.10	82.42	24.50	94.24	100
16	94.05	89.29	80.95	94.05	17.86	89.29	97.42
OA	81.67	91.93	92.36	93.61	70.47	94.52	95.63
AA	78.60	86.76	83.99	90.32	54.19	87.60	92.74
kappa	0.7902	0.9076	0.9124	0.9270	0.6545	0.9374	0.9512

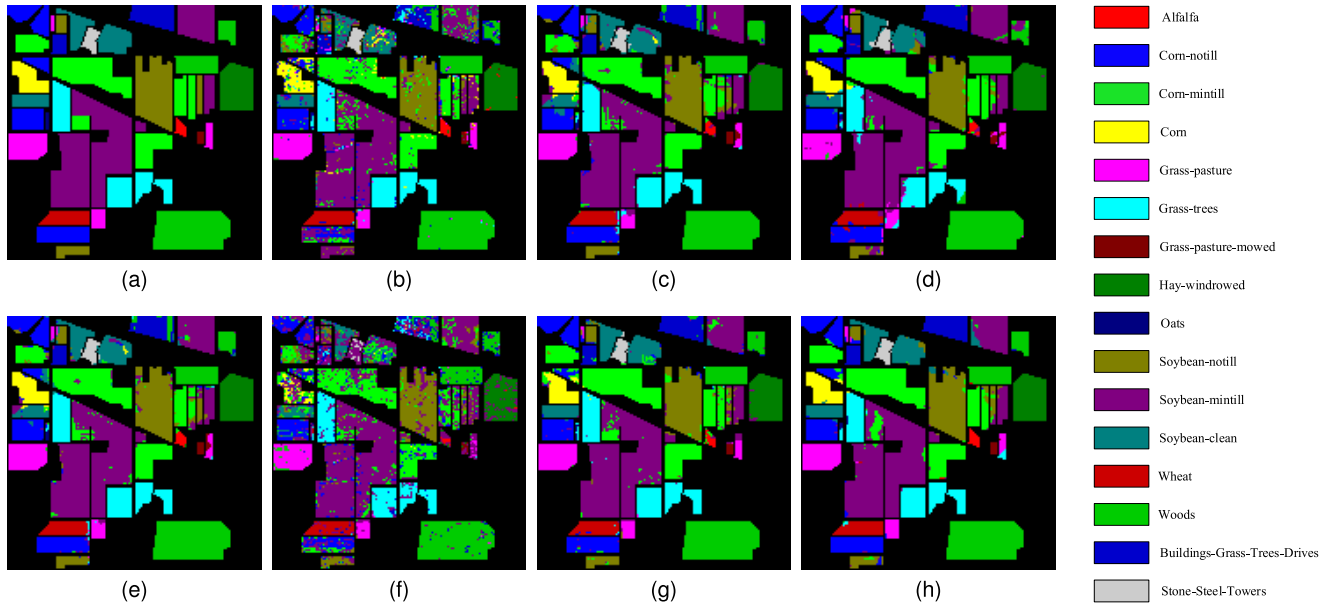


Fig. 2. Ground truth and classification maps for Indian Pines. (a) Ground truth. (b) SVM (81.67%). (c) SVMCK (91.93%). (d) JRSRC (92.36%). (e) cdSRC (93.61%). (f) LRR (70.47%). (g) LGIDL (94.52%). (h) LSLRR (95.63%).

kappa coefficient (κ), to evaluate the performance of different classification methods. Specifically, OA index denotes the percentage of HSI pixels that are classified correctly. AA index refers to the average value of accuracy of each class. However, both OA and AA indices only involve the errors of commission and they do not cover the user accuracy. The kappa coefficient (κ), a more reasonable measurement, involves not only the errors of commission but also the errors of omission.

C. Experimental Results and Analyses

1) *Indian Pines*: We randomly select 10% labeled samples in each class as the training set and the rest as the testing set.

Table I demonstrates the final classification performance (i.e., the accuracy for each category, OA, AA and kappa coefficient κ) for the Indian Pines data set. The corresponding classification maps of each algorithm are shown in Fig. 2. Among these comparison algorithms of HSI classification, SVM and LRR are pixelwise classification methods that only utilize the spectral feature. Other algorithms (SVMCK, JRSRC, LGIDL, and LSLRR) combine both spectral and spatial information to classify the HSI data. One can be seen easily from Table I that classification accuracy of SVM and LRR is far lower (OA decreases at least 10%) than that of the other methods. This indicates that the contextual feature can bring a great help for HSI classification. In addition, SVM

TABLE II
CLASSIFICATION ACCURACY (%) OF DIFFERENT COMPARISON METHODS AND THE PROPOSED LSLRR FOR PAVIA UNIVERSITY DATA SET

Class	SVM	SVMCK	JRSRC	cdSRC	LRR	LGIDL	LSLRR
1	94.11	96.02	95.62	96.57	88.46	96.81	97.33
2	96.94	99.63	99.26	99.44	97.06	99.79	99.98
3	81.44	82.40	88.82	89.27	72.67	89.22	91.98
4	94.37	97.32	91.69	93.27	74.41	98.18	98.73
5	99.30	97.03	99.84	99.92	68.47	100	100
6	86.73	95.63	94.91	96.19	67.54	99.35	99.90
7	86.30	89.47	87.89	92.64	80.36	94.70	96.52
8	84.02	91.14	92.68	94.08	82.68	92.17	94.97
9	99.89	98.00	99.59	99.89	94.56	98.89	99.11
OA	93.05	95.86	96.24	97.02	86.73	97.81	98.52
AA	91.46	94.07	94.47	95.70	80.69	96.57	97.61
kappa	0.9078	0.9524	0.9499	0.9605	0.8186	0.9710	0.9804

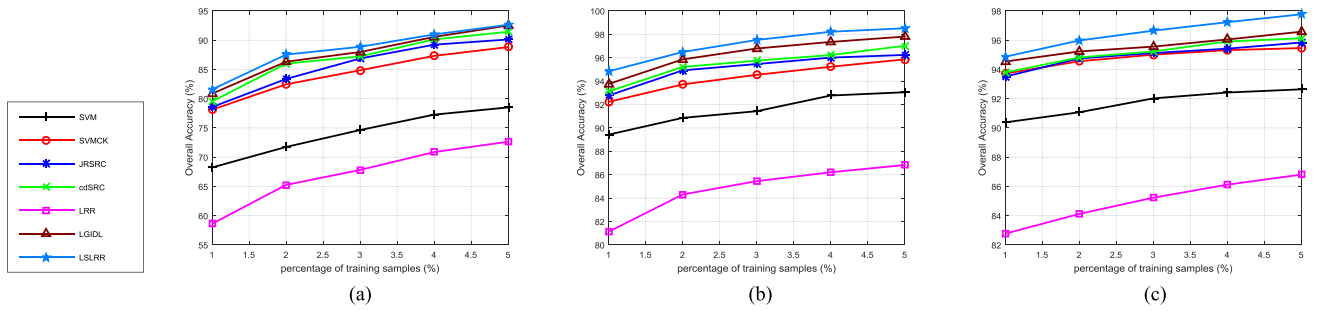


Fig. 3. Comparison of OA for all methods under different percentage of training samples. (a) Indian Pines. (b) Pavia University. (c) Salinas.

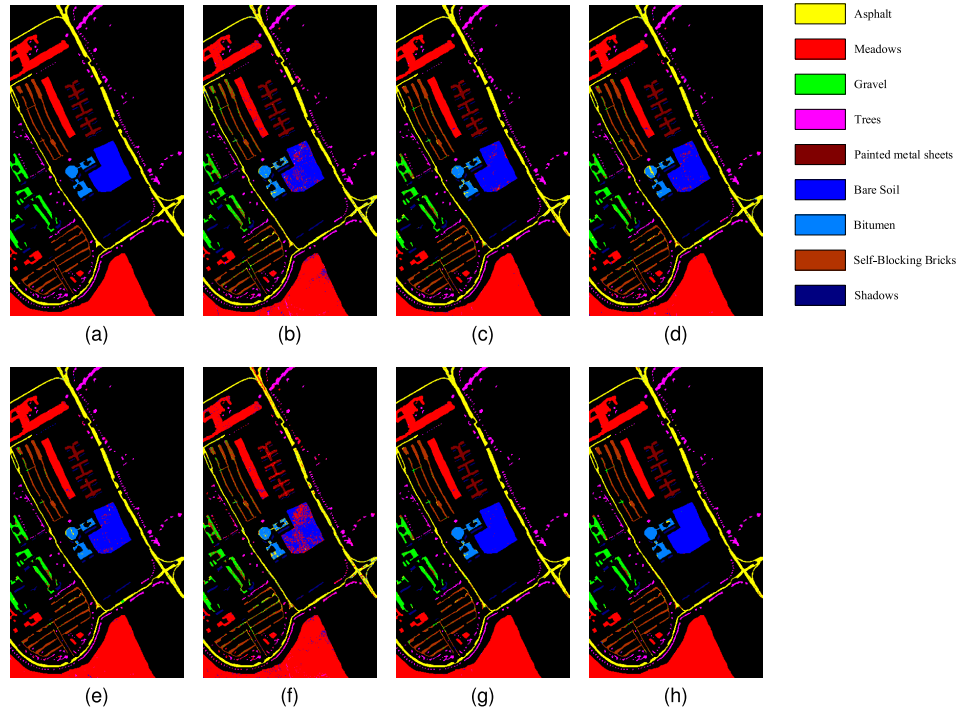


Fig. 4. Ground truth and classification maps for Indian Pines. (a) Ground truth. (b) SVM (81.67%). (c) SVMCK (91.93%). (d) JRSRC (92.36%). (e) cdSRC (93.61%). (f) LRR (70.47%). (g) LGIDL (94.52%). (h) LSLRR (95.63%).

outperforms the LRR a lot, which verifies the popular SVM is a superior classification algorithm. For the classification accuracy of every class in Table I, LGIDL achieves the best

result for the second class. JRSRC achieves the best result for the third class. cdSRC achieves the best result for the sixth class. The proposed LSLRR also obtains the highest accuracy

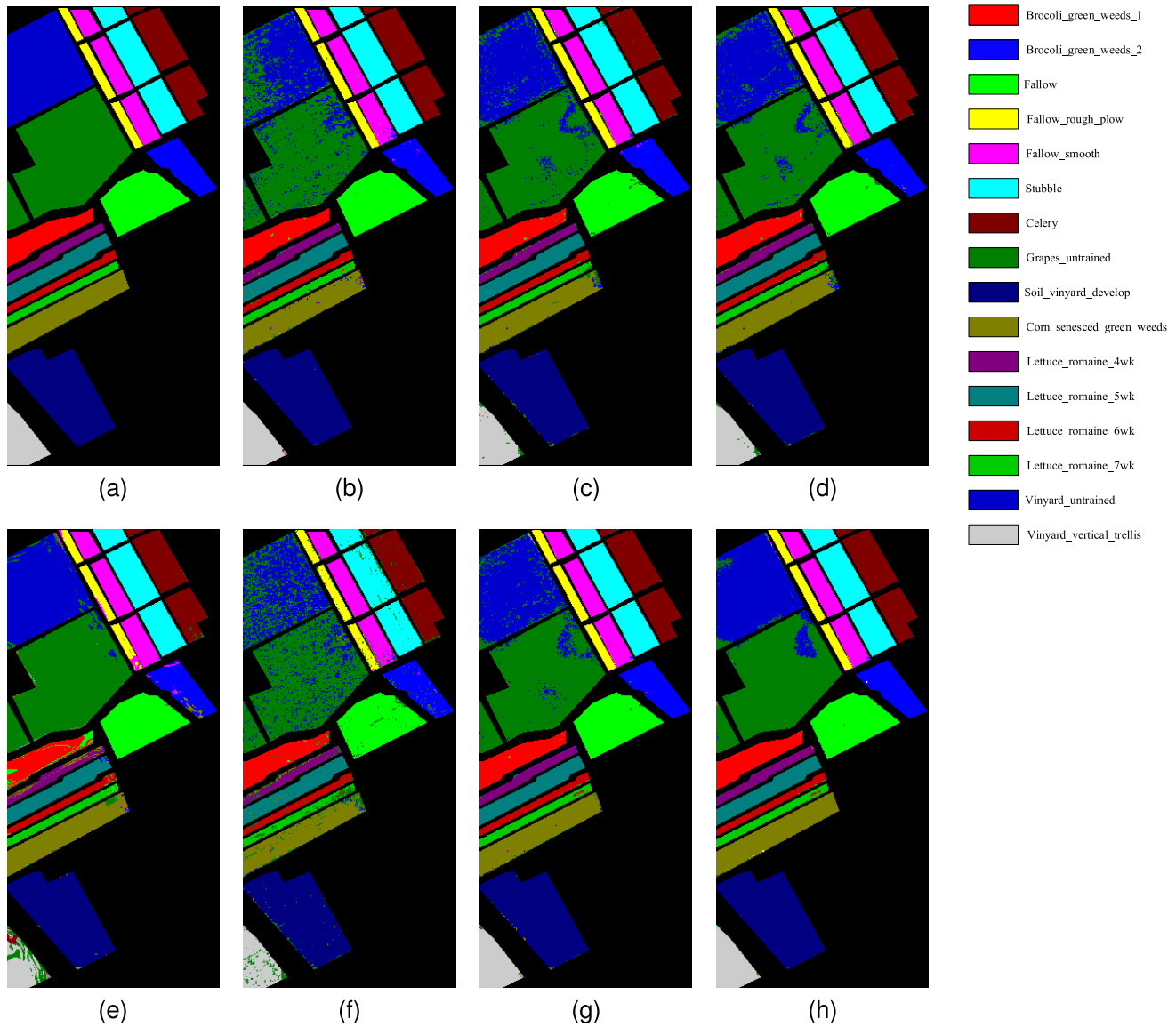


Fig. 5. Ground truth and classification maps for Indian Pines. (a) Ground truth. (b) SVM (81.67%). (c) SVMCK (91.93%). (d) JRSRC (92.36%). (e) cdSRC (93.61%). (f) LRR (70.47%). (g) LGIDL (94.52%). (h) LSLRR (95.63%).

in most classes. Furthermore, the classification OA of the proposed LSLRR improves more than 20% compared with the classical LRR. This is because LCC helps LRR in capturing the local feature, and SPS makes the solution \hat{Z} close to the ideal block-diagonal matrix. Moreover, Table I also obviously demonstrates that LSLRR has achieved the best performance than all other comparison methods. Fig. 3(a) illustrates the classification accuracy of various methods when a different number of samples are considered as the training set. It can be clearly observed that the classification performance of SVM and LRR is the worst, and other classification methods all have a promising performance. Among these, the proposed LSLRR yields the best classification results.

2) *Pavia University*: 5% of labeled HSI pixels are chosen to be the training set, and the remaining 95% are used for testing. In order to compare the experimental results quantitatively and visually, Table II and Fig. 4 exhibit the classification performance of Pavia University, and the corresponding visual

maps of all methods, respectively. As shown in Table II and Fig. 4, only a small number of HSI pixels are classified wrongly, and the classification accuracy of LSLRR is the highest in three evaluation indices. Except for the ninth class, LSLRR achieves the best results for other eight classes. This indicates that LSLRR is an effective and superior approach to classify HSIs. After incorporating the spatial characteristics into the composite kernels, SVMCK yields better classification results in almost all the classes compared with SVM. Similarly, the OA of original LRR is the lowest, and the main sample that is wrongly classified is class 3, 4, 5, and 6. As shown in Fig. 4(f), there are many red pixels (class 2) in the blue regions (class 6). Through improving LRR by two powerful techniques, LSLRR achieves OA of 99.9% in the sixth class. Compared with LRR, OA of LSLRR improves nearly 12%, and kappa coefficient (κ) of LSLRR improves more than 16%. Furthermore, we also investigate the influence of different number of training pixels on classification

TABLE III
CLASSIFICATION ACCURACY (%) OF DIFFERENT COMPARISON METHODS AND THE PROPOSED LSLRR FOR SALINAS DATA SET

Class	SVM	SVMCK	JRSRC	cdSRC	LRR	LGIDL	LSLRR
1	99.32	98.59	98.48	88.24	96.12	99.53	99.69
2	99.87	98.14	98.81	99.86	96.07	99.46	99.95
3	99.52	99.04	99.25	91.21	94.25	99.73	99.73
4	98.79	99.40	98.04	87.31	92.37	99.09	99.17
5	97.76	97.41	97.44	99.65	93.87	98.86	99.06
6	99.67	98.94	98.75	99.79	94.42	99.60	99.79
7	99.57	98.44	99.26	98.71	96.62	99.50	99.67
8	88.35	92.82	93.59	97.64	82.73	93.99	94.27
9	99.86	99.02	99.44	99.63	97.88	99.41	99.92
10	95.41	93.38	94.84	96.08	87.32	97.85	99.10
11	96.45	92.32	95.86	78.23	79.70	96.75	99.21
12	99.67	99.73	100	97.87	94.81	99.95	100
13	97.74	96.21	94.48	93.68	92.64	97.82	98.01
14	96.95	92.03	94.39	91.73	72.44	92.62	94.09
15	68.79	88.94	88.20	96.09	61.87	89.04	95.10
16	99.30	96.45	96.97	77.73	86.49	98.66	99.42
OA	92.64	95.46	95.84	96.13	86.81	96.58	97.77
AA	96.06	96.30	96.74	93.34	88.72	97.62	98.51
kappa	0.9179	0.9494	0.9536	0.9524	0.8520	0.9619	0.9752

accuracy for Pavia University set. And the corresponding figure is demonstrated in Fig. 3(b). Interestingly, the curve of LSLRR is the highest while that of LRR is the lowest, which reveals the improvement of LSLRR for LRR is successful.

3) *Salinas*: Similar to Pavia University, 5% pixels are selected to train classification model and the rest 95% are as the testing set. The classification accuracy of comparing methods and LSLRR is displayed in Table III. For the purpose of visualization, the classification maps are illustrated in Fig. 5. From the visual maps, the most wrongly classified pixels are in the dark blue (class 8) and dark green (class 15) regions. This is because the land surfaces of the eighth and the fifteenth classes have homologous properties, and the corresponding spectral reflectance curves are very similar. In addition, it is easy to observe that the proposed LSLRR yields the best accuracy compared with other methods, which justifies the effectiveness of LSLRR. From Table III, we can see that the classification accuracy of most classes is more than 99% and all OA is not lower than 94%. Moreover, Fig. 3(c) exhibits the OA of different methods for Salinas scene versus the percentage of training samples. This clearly displays that LSLRR can still obtain the best performance although a small number of pixels are used for the training set.

Fig. 6 exhibits the OA of three HSI data sets when the value of parameter m changes. Other crucial parameters are followed as Section V-B. m is an important parameter to control the weight of spatial information in the LCC. From Fig. 6, we can get that the optimal value of m is 25, 15, and 40 for Indian Pines, Pavia University, and Salinas, respectively. The method we employ to measure the spatial similarity is by Euclidean distance, which is more suitable for pixels of the same class distributing in a square or circular shape. As shown in Fig. 6, the shapes of many classes in Pavia University are slender,

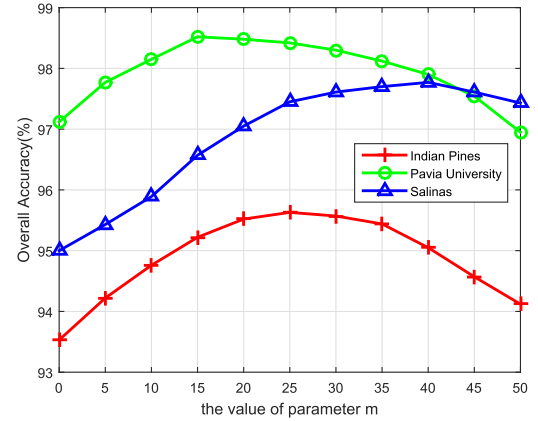


Fig. 6. OA of three HSI data sets under different values of parameter m .

while Salinas has many pixels whose distribution is more uniform. Therefore, the most appropriate m for Salinas is the largest, and that for Pavia University is the smallest. In summary, the large value of m is more reasonable for the HSI data set, which has higher compactness for each class.

Figs. 7 and 8 illustrate the OA of three HSI data sets under different values of parameter α and β , respectively. When investigating the influence of classification accuracy about parameter α or β , other parameters are set as the optimal values. Obviously, the optimal values of α for three data sets are 0.8, 0.3 and 1.0, respectively. When the LCC is not added, i.e., $\alpha = 0$, the classification accuracy decreases a lot comparing with the highest OA for all three data sets. Especially, for Indian Pines, OA decreases more than 12%. This indicates that LCC is extremely important for the proposed LSLRR. Furthermore, one can easily see that the optimal values of β for three data sets are 0.6, 1.2 and 0.4, respectively. Similarly, when $\beta = 0$, the classification accuracy is very low.

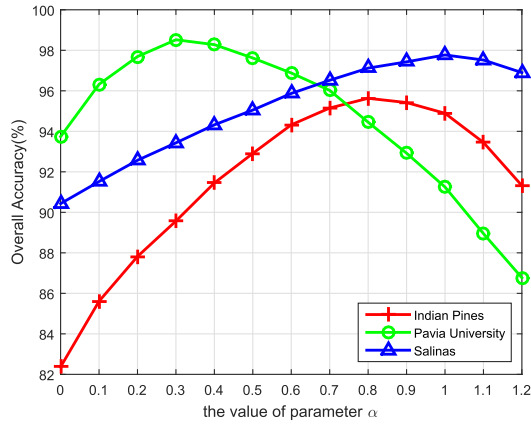


Fig. 7. OA of three HSI data sets under different values of parameter α .

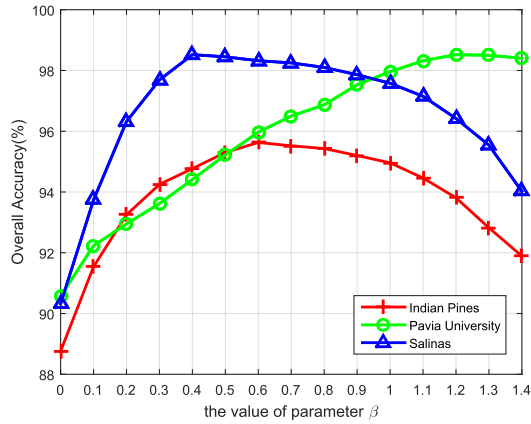


Fig. 8. OA of three HSI data sets under different values of parameter β .

TABLE IV
RUNNING TIME OF DIFFERENT HSI CLASSIFICATION METHODS

Methods	OA(%)	AA(%)	Kappa	Time(s)
SVM	81.67	78.60	0.7902	4.23
SVMCK	91.93	86.76	0.9076	6.17
JRSRC	92.36	83.99	0.9124	328.62
cdSRC	93.61	90.32	0.9270	118.86
LRR	70.47	54.19	0.6545	242.37
LGIDL	94.52	87.60	0.9374	382.13
LSLRR	95.63	92.74	0.9512	336.25

And the OA index improves rapidly when the value of β starts to increase from 0. It demonstrates the importance of SPS. To sum up, both LCC and SPS can provide a great deal of help to improve significantly the classification accuracy.

D. Comparison of Running Time

As follows, in order to testify the efficiency of the proposed LSLRR, we use running time to compare the computational complexity of all algorithms. Indian Pines data set is considered as an example, and 10% of labeled pixels of each class are used for training model. The experiments are conducted in MATLAB R2015a on a PC of Intel Core i7-3770 3.40 GHz CPU with 32-GB RAM. Table IV illustrates OA, AA, kappa

coefficient and running time of every method. According to the table, the time consumed by SVM and SVMCK are the least, but their classification accuracy is not high enough compared with JRSRC, cdSRC, LGIDL, and LSLRR. JRSRC and LGIDL can obtain promising classification performance, but the running time is too long. For the proposed LSLRR, it is computationally acceptable and the classification accuracy is the highest.

VI. CONCLUSION

In this paper, a novel LSLRR is proposed to classify HSIs. In order to overcome the drawbacks of traditional LRR, LSLRR introduces two key techniques, LCC and SPS, to improve LRR and make it more suitable for HSI classification. In LSLRR, a new similarity metric combining both spatial and spectral characteristics is first presented. Then, LCC utilizes the new similarity metric to make HSI pixels with large distance have a small similarity, which can easily capture the local structure. In addition, SPS makes the solution of LSLRR close to a classwise block-diagonal matrix. Finally, the classification results can be easily obtained without any complex classifiers. Extensive experiments on three public HSI data sets are carried out to evaluate the performance of the proposed LSLRR, and the experimental results show that LSLRR outperforms other state-of-the-art comparison methods.

REFERENCES

- [1] Q. Wang, Z. Meng, and X. Li, "Locality adaptive discriminant analysis for spectral-spatial classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 2077–2081, Nov. 2017.
- [2] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1, pp. 55–63, Jan. 1968.
- [3] G. Camps-Valls, L. Gomez-Chova, J. Muñoz-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.
- [4] 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.
- [5] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 663–670.
- [6] Y. Li, J. Liu, Z. Li, Y. Zhang, H. Lu, and S. Ma, "Learning low-rank representations with classwise block-diagonal structure for robust face recognition," in *Proc. AAAI*, 2014, pp. 2810–2816.
- [7] L. Li, S. Li, and Y. Fu, "Learning low-rank and discriminative dictionary for image classification," *Image Vis. Comput.*, vol. 32, no. 10, pp. 814–823, 2014.
- [8] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.
- [9] X. Zhou, C. Yang, and W. Yu, "Moving object detection by detecting contiguous outliers in the low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 597–610, Mar. 2013.
- [10] Q. Wang, Z. Yuan, and X. Li, "GETNET: A general end-to-end two-dimensional CNN framework for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2018.2849692.
- [11] X. Sun, Q. Qu, N. M. Nasrabadi, and T. D. Tran, "Structured priors for sparse-representation-based hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 7, pp. 1235–1239, Dec. 2013.
- [12] S. Mei, Q. Bi, J. Ji, J. Hou, and Q. Du, "Spectral variation alleviation by low-rank matrix approximation for hyperspectral image analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 6, pp. 796–800, Jun. 2016.

- [13] Q. Wang, J. Lin, and Y. Yuan, "Salient band selection for hyperspectral image classification via manifold ranking," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 6, pp. 1279–1289, Jun. 2016.
- [14] 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.
- [15] U. Srinivas, Y. Chen, V. Monga, N. M. Nasrabadi, and T. D. Tran, "Exploiting sparsity in hyperspectral image classification via graphical models," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 3, pp. 505–509, May 2013.
- [16] 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.
- [17] C. Li, Y. Ma, X. Mei, C. Liu, and J. Ma, "Hyperspectral image classification with robust sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 13, no. 5, pp. 641–645, Mar. 2016.
- [18] J. Li, H. Zhang, and L. Zhang, "Efficient superpixel-level multi-task joint sparse representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5338–5351, Oct. 2015.
- [19] W. Fu, S. Li, L. Fang, and J. A. Benediktsson, "Adaptive spectral-spatial compression of hyperspectral image with sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 671–682, Feb. 2017.
- [20] L. Gan, J. Xia, P. Du, and J. Chanussot, "Multiple feature kernel sparse representation classifier for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: [10.1109/TGRS.2018.2814781](https://doi.org/10.1109/TGRS.2018.2814781).
- [21] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral-spatial hyperspectral image classification via multiscale adaptive sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7738–7749, Dec. 2014.
- [22] Y. Y. Tang and H. Yuan, "Manifold-based sparse representation for hyperspectral image classification," in *Handbook of Pattern Recognition and Computer Vision*. Singapore: World Scientific, 2016, pp. 331–350.
- [23] W. Li and Q. Du, "A survey on representation-based classification and detection in hyperspectral remote sensing imagery," *Pattern Recognit. Lett.*, vol. 83, no. 2, pp. 115–123, Nov. 2015.
- [24] Q. Wang, F. Zhang, and X. Li, "Optimal clustering framework for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: [10.1109/TGRS.2018.2828161](https://doi.org/10.1109/TGRS.2018.2828161).
- [25] Q. Qu, N. M. Nasrabadi, and T. D. Tran, "Abundance estimation for bilinear mixture models via joint sparse and low-rank representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4404–4423, Jul. 2014.
- [26] Y.-Q. Zhao and J. Yang, "Hyperspectral image denoising via sparse representation and low-rank constraint," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 296–308, Jan. 2015.
- [27] Q. Shi, B. Du, and L. Zhang, "Domain adaptation for remote sensing image classification: A low-rank reconstruction and instance weighting label propagation inspired algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5677–5689, Oct. 2015.
- [28] Y. Yuan, M. Fu, and X. Lu, "Low-rank representation for 3D hyperspectral images analysis from map perspective," *Signal Process.*, vol. 112, pp. 27–33, Jul. 2015.
- [29] A. Sumarsono and Q. Du, "Low-rank subspace representation for supervised and unsupervised classification of hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4188–4195, Sep. 2016.
- [30] A. Soltani-Farani, H. R. Rabiee, and S. A. Hosseini, "Spatial-aware dictionary learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 527–541, Jan. 2015.
- [31] Z. He, L. Liu, R. Deng, and Y. Shen, "Low-rank group inspired dictionary learning for hyperspectral image classification," *Signal Process.*, vol. 120, pp. 209–221, Mar. 2016.
- [32] S. Jia, X. Zhang, and Q. Li, "Spectral-spatial hyperspectral image classification using $\ell_{1/2}$ regularized low-rank representation and sparse representation-based graph cuts," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2473–2484, Jun. 2015.
- [33] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proc. IEEE*, vol. 98, no. 6, pp. 1045–1057, Jun. 2010.
- [34] K. Egan, S. O. Aase, and J. Hakon Husoy, "Method of optimal directions for frame design," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 5, Mar. 1999, pp. 2443–2446.
- [35] S. Lesage, R. Gribonval, F. Bimbot, and L. Benaroya, "Learning unions of orthonormal bases with thresholded singular value decomposition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 5, Mar. 2005, pp. v/293–v/296.
- [36] 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.
- [37] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
- [38] Z. Wang, N. M. Nasrabadi, and T. S. Huang, "Spatial-spectral classification of hyperspectral images using discriminative dictionary designed by learning vector quantization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4808–4822, Aug. 2014.
- [39] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.
- [40] A. Chakrabarti and T. Zickler, "Statistics of real-world hyperspectral images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 193–200.
- [41] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 612–620.
- [42] Z. Lin, M. Chen, and Y. Ma, (2010). "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices." [Online]. Available: <https://arxiv.org/abs/1009.5055>
- [43] K. Tang, R. Liu, Z. Su, and J. Zhang, "Structure-constrained low-rank representation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 12, pp. 2167–2179, Dec. 2014.
- [44] *Hyperspectral Remote Sensing Scenes*. Accessed: Jun. 1, 2018. [Online]. Available: http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes
- [45] Y. Xiao, H. Wang, and W. Xu, "Parameter selection of Gaussian kernel for one-class SVM," *IEEE Trans. Cybern.*, vol. 45, no. 5, pp. 941–953, May 2015.
- [46] M. Cui and S. Prasad, "Class-dependent sparse representation classifier for robust hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2683–2695, May 2015.



Qi Wang (M'15–SM'15) received the B.E. degree in automation and the Ph.D. degree in pattern recognition and intelligent systems from the University of Science and Technology of China, Hefei, China, in 2005 and 2010, respectively.

He is currently a Professor with the School of Computer Science, the Unmanned System Research Institute, and the Center for OPTical IMagery Analysis and Learning, Northwestern Polytechnical University, Xi'an, China. His research interests include computer vision and pattern recognition.



Xiang He received the B.E. degree in automation from Northwestern Polytechnical University, Xi'an, China, in 2017. He is currently pursuing the M.S. degree in computer science with the Center for Optical Imagery Analysis and Learning, Northwestern Polytechnical University.

His research interests include hyperspectral image processing and computer vision.

Xuelong Li (M'02–SM'07–F'12) is a Full Professor with the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China.