Hyperspectral Image Classification by Nonlocal Joint Collaborative Representation With a Locally Adaptive Dictionary

Jiayi Li, *Student Member, IEEE*, Hongyan Zhang, *Member, IEEE*, Yuancheng Huang, and Liangpei Zhang, *Senior Member, IEEE*

Abstract—Sparse representation has been widely used in image classification. Sparsity-based algorithms are, however, known to be time consuming. Meanwhile, recent work has shown that it is the collaborative representation (CR) rather than the sparsity constraint that determines the performance of the algorithm. We therefore propose a nonlocal joint CR classification method with a locally adaptive dictionary (NJCRC-LAD) for hyperspectral image (HSI) classification. This paper focuses on the working mechanism of CR and builds the joint collaboration model (JCM). The joint-signal matrix is constructed with the nonlocal pixels of the test pixel. A subdictionary is utilized, which is adaptive to the nonlocal signal matrix instead of the entire dictionary. The proposed NJCRC-LAD method is tested on three HSIs, and the experimental results suggest that the proposed algorithm outperforms the corresponding sparsity-based algorithms and the classical support vector machine hyperspectral classifier.

Index Terms—Classification, hyperspectral imagery, joint collaboration model, k-nearest neighbor (K-NN), sparse representation.

I. INTRODUCTION

HYPERSPECTRAL imaging spectrometer data, which represent a function of the wavelength with a large spectral range and a high spectral resolution, can facilitate the superior discrimination of land cover types [1], [2]. In supervised classification [3], [4], the label of each test pixel is determined by the corresponding training samples from each class. The support vector machine (SVM) classifier [5], which aims to find an optimal separating hyperplane between two classes to solve the binary classification problem, and some variations of the SVM-based algorithms, such as relevance vector machines (RVMs) [6], have been proved to be efficient tools for dealing with the classification problems of high-dimensional data

and have shown excellent performance in hyperspectral image (HSI) classification.

In the last few years, sparse-representation-based classification (SRC) [7], [8], as a linear regression approach with sparse-norm regularization, has been widely used in various pattern recognition applications and computer vision areas [9]–[13]. Several variants based on SRC [14]–[17] have also been implemented in HSI classification and have shown excellent performances. Chen *et al.* [14] utilized the spectral and contextual information of the HSI and proposed a joint sparse representation classification (JSRC) method under the assumption of a joint sparsity model (JSM) [18]. Sami ul Haq *et al.* [16] introduced a fast and robust sparse approach for HSI classification using just a few labeled samples. Qian *et al.* [17] proposed a hyperspectral feature extraction and pixel classification method based on structured sparse logistic regression and 3-D discrete wavelet transform texture features. Although these sparsity-based approaches can give better performances than the conventional HSI classifiers [19], [20], the computation costs for the methods mentioned earlier are quite high as the sparsity norm is a complicated procedure in optimization computation.

Zhang *et al.* [21] further revealed that it is not necessary to regularize the sparse coefficient α by the computationally expensive ℓ_1-norm regularization when the feature dimension is high enough. They pointed out that it is the collaborative representation (CR), which represents the test pixel collaboratively with training samples from all the classes, rather than the sparsity constraint, that determines the performance of the algorithm and proposed several specific classifiers based on CR for face recognition. Since the hyperspectral pixel denoted by a vector is also a high-dimensional signal, the CR classification with regularized least squares (RLS) (referred to as CRC in this paper) scheme may be more suitable for HSI processing. CRC implements the CR by an ℓ_2-norm-regularized linear regression scheme with much lower computational cost. Compared with the sparsity-based classification algorithms, CRC can achieve a very competitive accuracy and has a significantly lower complexity.

In this paper, we focus on the CR working mechanism and propose a novel Frobenius-norm linear-regression-constrained nonlocal joint CR classification method with a locally adaptive dictionary (NJCRC-LAD) for hyperspectral imagery. The contributions of this paper are in three aspects. First, we propose a general joint collaboration model inspired by the JSM to

incorporate the neighborhood information. In hyperspectral imagery, the pixels in a small patch often consist of materials with similar characteristics, which can be simultaneously linearly represented by the same feature subspace. Second, a nonlocal joint-signal matrix construction method is introduced, with the consideration of rejecting the dissimilar neighboring pixels. As the superiority of the simultaneous linear representation will work if and only if the signal matrix fulfills the assumption of sharing a commonly represented pattern, as introduced in [22], we select the most similar K spectrally correlated pixels in a $\sqrt{T} \times \sqrt{T}$ spatial patch and construct a nonlocal joint-signal matrix by stacking these pixels. Finally, a locally adaptive dictionary construction approach is presented. The pixel to be tested is collaboratively represented over the locally adaptive dictionary without the whole dictionary. In this paper, we make use of k-nearest neighbor (K-NN) [23]–[26] to constrain the discriminated projections, and then require that the test sample should be represented by these selected training samples. To sum up these contributions, the proposed NJCRC-LAD method aims at realizing the pixels with a nonlocal self-similarity to be simultaneously reconstructed in locally collaborative learning, thereby acquiring an improved HSI classification performance with a reasonable computational burden.

We now introduce the relationship between the proposed NJCRC-LAD method and the other conventional hyperspectral classifiers. Although SVM and RVM also make use of sparse support vectors/sparse relative vectors, the CR-based classifier is different from these conventional hyperspectral classifiers (multinomial logistic regression, SVM, and its extension RVM) in the following aspects. First, the SVM and its variation benchmarks are all binary classifiers, while the CR-based classifier works from a reconstruction point of view and labels the pixel from the multiple classes directly. Second, when a new training sample joins the training set, the dictionary of the proposed algorithm can easily contain this sample without retraining the model, whereas the other classifiers need to retrain the model for the new training data. Last but not the least, the conventional classifiers have an explicit and time-consuming training stage and are trained only once to obtain the fixed parameters, which will be used to classify all of the test data. On the other hand, with our proposed CR-based approach, a CR vector for each test pixel is estimated, utilizing a set of specific locally adaptive training samples from the entire training data set. That is to say, the conventional classifiers will cost more time in the training stage, while the dominant cost for the NJCRC-LAD classifier comes from the representation estimation stage.

The remaining parts of this paper are organized as follows. Section II briefly introduces the CR mechanism and two standard specific classifiers (SRC and CRC). Section III proposes the NJCRC-LAD algorithm for hyperspectral imagery. The experimental results of the proposed algorithm are given in Section IV. Finally, Section V concludes this paper.

II. CLASSIFICATION OF HSI VIA CR

In this section, we first introduce the general model of CR in hyperspectral imagery by reconstructing the test pixel by a linear combination of training samples over the global dictionary. We then describe the existing classifiers (i.e., SRC [7], [8], and CRC [21]) and their corresponding regularizations.

A. General CR Model

In the general CR model, the spectral signature of the test pixel is approximated by the training samples from all the classes. Every hyperspectral pixel can be denoted as a B-dimensional vector, where B refers to the number of bands of the HSI. Suppose that we have M distinct classes and stack the given N_i $(i = 1, \ldots, M)$ training pixels from the ith class as columns of a subdictionary $\mathbf{A}_i = [\boldsymbol{a}_{i,1}, \boldsymbol{a}_{i,2}, \ldots, \boldsymbol{a}_{i,N_i}] \in \mathbb{R}^{B \times N_i}$; then, the collaborative dictionary $\mathbf{A} \in \mathbb{R}^{B \times N}$ with $N = \sum_i^M N_i$ is constructed by combining all the subdictionaries $\{\mathbf{A}_i\}_{i=1,\ldots,M}$. In the CR model, the hyperspectral signal $\boldsymbol{s}$ which belongs to the ith class can be collaboratively represented as a linear combination of all the given training samples

$$\boldsymbol{s} = \mathbf{A}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} = \mathbf{A}_1\boldsymbol{\alpha}_1 + \cdots + \mathbf{A}_i\boldsymbol{\alpha}_i + \cdots + \mathbf{A}_M\boldsymbol{\alpha}_M + \boldsymbol{\varepsilon}$$

$$= \mathbf{A}_i\boldsymbol{\alpha}_i + \sum_{j=1, j \neq i}^{M} \mathbf{A}_j\boldsymbol{\alpha}_j + \boldsymbol{\varepsilon} \in \mathbb{R}^B \qquad (1)$$

where the whole space constitutes a dominant low-dimensional subspace spanned by $\mathbf{A}_i$ and a complementary subspace spanned by the rest of the training samples, which can be considered as an external collaborative partner to the dominant subspace. The vector $\boldsymbol{\alpha}$ associated with the whole dictionary can be denoted as a general CR coefficient. For a signal $\boldsymbol{s}$ from class i, the simplest method to find its label is by the least squares method, which can be expressed as $\text{class}(\boldsymbol{s}) = \arg\min_i \|\boldsymbol{s} - \mathbf{A}_i\boldsymbol{\alpha}_i\|_2^2$. However, in practice, a pixel with some light disturbance may be misclassified to class j $(j \neq i)$ as $\|\boldsymbol{s} - \mathbf{A}_i\boldsymbol{\alpha}_i\|_2^2 > \|\boldsymbol{s} - \mathbf{A}_j\boldsymbol{\alpha}_j\|_2^2$, which leads to an unstable classification result [21]. This problem can be much alleviated by regularization. A general model of CR can be represented as

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{s} - \mathbf{A}\boldsymbol{\alpha}\|_q \qquad \text{s.t. } \|\boldsymbol{\alpha}\|_p \leq \varepsilon \qquad (2)$$

where ε is a small threshold and p and q are equal to one or two. Different settings of p and q lead to different instantiations of the CR model. In theory, p and q represent the distribution of the image noise and the prior of the coefficient, the range of p is $[0, \infty)$, and that of q is $(0, \infty)$.

B. Reconstruction and Classification of HSI via CR

In the well-known SRC scheme [7], [8], p is set as one while q is set as one or two to handle the face recognition problem, with or without occlusion/corruption, respectively. For hyperspectral imagery with Gaussian noise, we set q as two, and the Lagrangian dual form of this case can be shown as

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \{\|\boldsymbol{s} - \mathbf{A}\boldsymbol{\alpha}\|_2 + \lambda\|\boldsymbol{\alpha}\|_1\} \qquad (3)$$

where the parameter λ is a tradeoff between the data fidelity term and the coefficient prior. The ℓ_1-norm regularization [27]–[30] in this formulation suggests that the signal should be classified to the class which can faithfully represent it using less

and nonlocal similar training samples [21]. λ makes a tradeoff between the sparsity constraint term and the data fidelity term. The class of signal s can be determined by minimizing the residual r_i (i.e., the error between s and the linearly recovered approximated result from the training samples in the ith class)

$$\text{class}(\boldsymbol{s}) = \arg \min_{i=1,\ldots,M} r_i(\boldsymbol{s})$$
$$= \arg \min_{i=1,\ldots,M} \|\boldsymbol{s} - \mathbf{A}_i \hat{\boldsymbol{\alpha}}_i\|_2. \quad (4)$$

While this classifier can achieve an excellent classification performance, it should be noted that the ℓ_1-norm optimization problem is quite time consuming because of the computational complexity caused by the nonsmoothing constraint. For classification, Zhang *et al.* [21] suggested that the nonsparse ℓ_2-norm regularization could play a similar role to the sparse norm in enhancing the discrimination of representation. In order to reduce the complexity of the algorithm and impose the role of CR, the ℓ_1-norm sparsity constraint is replaced by the ℓ_2-norm "sparsity constraint" [21] in the CRC scheme

$$\hat{\boldsymbol{\alpha}} = \arg \min_{\boldsymbol{\alpha}} \{\|\boldsymbol{s} - \mathbf{A}\boldsymbol{\alpha}\|_2 + \lambda\|\boldsymbol{\alpha}\|_2\} \quad (5)$$

where λ is the regularization parameter. This regularization term in (5) not only ensures that the least squares solution is stable but also introduces a weaker "sparsity-based" constraint than the ℓ_1-norm regularization. For classification, where $\hat{\boldsymbol{\alpha}}_i$ is the coding vector associated with class i, the ℓ_2-norm $\|\hat{\boldsymbol{\alpha}}_i\|_2$ also brings some discriminative information. The classification rule for CRC via RLS is denoted as

$$\text{class}(\boldsymbol{s}) = \arg \min_{i=1,\ldots,M} \|\boldsymbol{s} - \mathbf{A}_i \hat{\boldsymbol{\alpha}}_i\|_2 / \|\hat{\boldsymbol{\alpha}}_i\|_2. \quad (6)$$

III. NONLOCAL JCRC WITH A LOCALLY ADAPTIVE CONSTRAINT

This section is organized as follows. We first build the general joint CR model and point out that the JSM and JCM are both specific instantiations of the general model. Second, we incorporate nonlocal spatial neighborhood information in the joint CR classifier by constructing a nonlocal joint-signal matrix, which consists of the highly correlated pixels in the neighboring window of the test one. Next, we restrict the global training samples to construct a locally adaptive dictionary which caters to the CR in the hyperspectral feature space. Finally, we integrate the individual modules described previously into a supervised NJCRC-LAD for hyperspectral imagery.

A. General Joint CR Model and Joint CR Classification

Inspired by the JSM [18] used in HSI classification [14] and its superior performance, we build the general joint CR model by extending the vector-oriented general CR model to the 2-D matrix case, to improve the representation result by exploiting the spatial correlation across neighboring pixels. In this model, we consider that the HSI pixels with high spectral similarities are approximated by a linear combination of common atoms

from a given structured dictionary. That is, the corresponding dominant subspace for these pixels should be the same, and the atoms in the subspace are weighted with a different set of coefficients for each pixel.

We next illustrate this model with an HSI patch, where pixels are assumed to consist of similar materials. We denote $\boldsymbol{s}_p$ as the spectral signal of one pixel p in this patch. With a given $B \times N$ structured dictionary $\mathbf{A}$, $\boldsymbol{s}_p$ can be linearly represented as

$$\boldsymbol{s}_p = \mathbf{A}\boldsymbol{\alpha}_p + \boldsymbol{\varepsilon}_p = \mathbf{A}_i\boldsymbol{\alpha}_{i,p} + \sum_{j=1,j\neq i}^{M} \mathbf{A}_j\boldsymbol{\alpha}_{j,p} + \boldsymbol{\varepsilon}_p \quad (7)$$

where $\mathbf{A}_i$ denotes the low-dimensional dominant subspace and $\boldsymbol{\varepsilon}_p$ is a random noise vector. As another pixel q located in the same patch consists of similar materials to $\boldsymbol{s}_p$, it can also be approximated by a linear combination of the same dominant subspace and its collaborative partner

$$\boldsymbol{s}_q = \mathbf{A}\boldsymbol{\alpha}_q + \boldsymbol{\varepsilon}_q = \mathbf{A}_i\boldsymbol{\alpha}_{i,q} + \sum_{j=1,j\neq i}^{M} \mathbf{A}_j\boldsymbol{\alpha}_{j,q} + \boldsymbol{\varepsilon}_q \quad (8)$$

where pixel q corresponds to the same dominant subspace $\mathbf{A}_i$ with pixel p.

Let the neighboring window size be set as T, and the HSI patch is stacked to construct the joint-signal matrix $\mathbf{S} = [\boldsymbol{s}_1 \ \boldsymbol{s}_2 \ \cdots \ \boldsymbol{s}_T]$ with the size of $B \times T$. Using the joint collaboration model, $\mathbf{S}$ can be represented by

$$\mathbf{S} = [\boldsymbol{s}_1 \ \boldsymbol{s}_2 \ \cdots \ \boldsymbol{s}_T] = [\mathbf{A}\boldsymbol{\alpha}_1 + \boldsymbol{\varepsilon}_1 \ \mathbf{A}\boldsymbol{\alpha}_2 + \boldsymbol{\varepsilon}_2 \ \cdots \ \mathbf{A}\boldsymbol{\alpha}_T + \boldsymbol{\varepsilon}_T]$$
$$= \mathbf{A}\underbrace{[\boldsymbol{\alpha}_1 \ \boldsymbol{\alpha}_2 \ \cdots \ \boldsymbol{\alpha}_T]}_{\boldsymbol{\Psi}} + \boldsymbol{\Sigma}$$
$$= \mathbf{A}_i\boldsymbol{\Psi}_i + \sum_{j=1,j\neq i}^{M} \mathbf{A}_j\boldsymbol{\Psi}_j + \boldsymbol{\Sigma} \quad (9)$$

where $\boldsymbol{\Psi}$ is a set of all the coefficient vectors $\{\boldsymbol{\alpha}_t\}_{t=1,\ldots,T}$ and $\boldsymbol{\Psi}_i$ is the subset of $\boldsymbol{\Psi}$. It is assumed that all the neighboring pixels share the same low-dimensional dominant subspace with different coefficients. $\boldsymbol{\Sigma}$ is the model noise matrix corresponding to the joint-signal matrix.

Corresponding to the general CR model, the general joint CR model can be represented as

$$\hat{\boldsymbol{\Psi}} = \arg \min_{\boldsymbol{\Psi}} \{\|\mathbf{S} - \mathbf{A}\boldsymbol{\Psi}\|_p\} \qquad \text{s.t. prior}(\boldsymbol{\Psi}) \quad (10)$$

where different p's and priors of the coefficient matrix lead to different instantiations of this general model. When setting p as the Frobenius norm and the sparse row prior constraint to the coefficient matrix, this model tends to the JSM in [14].

In hyperspectral imagery, by assuming that the scene contains only Gaussian noise, the joint-signal matrix $\mathbf{S}$ can be reconstructed by solving the following joint collaborative recovery via a Frobenius-norm optimization problem:

$$\hat{\boldsymbol{\Psi}} = \arg \min_{\boldsymbol{\Psi}} \left\{\|\mathbf{S} - \mathbf{A}\boldsymbol{\Psi}\|_F^2 + \lambda\|\boldsymbol{\Psi}\|_F^2\right\}. \quad (11)$$

Corresponding to the JSM, this equation is referred to as the JCM. The solution of the joint collaborative coding with RLS

in (11) can be easily and analytically derived as

$$\hat{\mathbf{\Psi}} = (\mathbf{A}'\mathbf{A} + \lambda \cdot \mathbf{I})^{-1}\mathbf{A}' \cdot \mathbf{S} \tag{12}$$

where $\mathbf{A}'$ is the transpose of $\mathbf{A}$. It should be noted that, from the perspective of mathematical computation, the coefficient matrix $\hat{\mathbf{\Psi}}$ can be obtained by separately stacking the coefficient vectors $(\mathbf{A}'\mathbf{A} + \lambda \cdot \mathbf{I})^{-1}\mathbf{A}' \cdot \mathbf{S}$ for every pixel $\boldsymbol{s}_i$ in $\mathbf{S}$. For consistence and convenience, the induced classification method is still referred to as JCRC. Once the coefficient matrix $\hat{\mathbf{\Psi}}$ is obtained, analogous to (6), the classification rule of the JCRC method is denoted as

$$\text{class}(\boldsymbol{s}) = \arg \min_{i=1,\dots,M} \|\mathbf{S} - \mathbf{A}_i\hat{\mathbf{\Psi}}_i\|_F^2 / \|\hat{\mathbf{\Psi}}_i\|_F^2. \tag{13}$$

B. Nonlocal Joint-Signal Matrix

While similar pixels tend to locate in an image spatial patch, there will still be neighborhood pixels with low correlation in heterogeneous areas, particularly around the image edges. It is not a good idea to roughly construct the joint-signal matrix by picking all the available spatial neighboring pixels around the central test one. In this section, we introduce a nonlocal joint-signal matrix construction approach to reject the dissimilar neighboring pixels.

We select a large neighborhood window centered at test pixel $\boldsymbol{s}_c$, with a size of $\sqrt{T} \times \sqrt{T}$. We consider that only $K - 1$ ($K \le T - 1$) neighboring pixels are similar to the test pixel and should be stacked into the nonlocal joint-signal matrix $\mathbf{S}_K$, and the other pixels in the neighborhood window are discarded. The K-NN method is used to construct the joint-signal submatrix by

$$\boldsymbol{simi} = \mathbf{S}' \cdot \boldsymbol{s}_c \tag{14}$$

where $\mathbf{S}'$ denotes the transpose of the neighboring signal set $\mathbf{S}$ which is sized as T and $\boldsymbol{s}_c$ is the test pixel. First, we calculate the correlation between every pixel in the neighborhood window and the test pixel $\boldsymbol{s}_c$ and get the correlation vectors $\boldsymbol{simi}$. We then sort the correlation coefficient vectors $\boldsymbol{simi}$ in a descending order and select the first K most correlated signals $\{\boldsymbol{s}_k\}_{k=1,\dots,K}$ to the test pixel $\boldsymbol{s}_c$ from $\mathbf{S}$. It is obvious that the first one is the test pixel $\boldsymbol{s}_c$ itself, and the rest of the $K - 1$ pixels are the ones which are the most correlated pixels with the central test pixel $\boldsymbol{s}_c$. It is believed that the K pixels $\{\boldsymbol{s}_k\}_{k=1,\dots,K}$ share a "common collaboration pattern" (the corresponding signal matrix is called the nonlocal joint-signal matrix in this paper) as they are selected by the measure of the correlations, not the spatial distance.

C. Locally Adaptive Dictionary

The working mechanism of CR is that some samples from other classes can be useful to represent the test pixel when training samples belonging to different classes share certain similarities [21]. In this way, atoms in the dictionary should be very similar in some certain feature pattern to each other in order to faithfully represent the test sample. However, one notable fact is that the spectral signals of different classes share

certain similarities as well as some differences. The fact that the spectral characteristics of several classes share similarities means that some samples from class j may be useful to represent the test pixel with label i. Meanwhile, it is also natural that the samples from class k may have a negative effect on the coding of the test pixel with label i when class i is quite different from class k. To the best of our knowledge, most CR-based classification methods [21], [31]–[33] utilize the set of all the high-dimensional training samples as the global dictionary, which can be considered as a highly redundant dictionary for a specific test pixel. In the hyperspectral classification case, for each test pixel, such a "general" dictionary is optimal when many training samples in this highly redundant dictionary are irrelevant to the test pixel. That is to say, these irrelevant samples may reduce the representation accuracy. Furthermore, as suggested in [23] and [34], locality is more important than sparsity, as locality must lead to sparsity, but not necessarily vice versa. We therefore propose to utilize an adaptive subdictionary selection strategy constrained by locality for the CR. The superior effectiveness of the adaptive subdictionary has been discussed in [35] for image restoration, and we here investigate the performance in HSI classification.

As described in the previous section, for the hyperspectral pixel $\boldsymbol{s}_c$ to be tested, we construct the nonlocal joint hyperspectral signal matrix $\mathbf{S}_K$. Our next task is to pick L training samples from all the training samples and construct the adaptive subdictionary for the current hyperspectral test pixel $\boldsymbol{s}_c$. Again, we use the K-NN method as the subdictionary selection scheme. First, we normalize each column of the general dictionary $\mathbf{A}$ and that of the $\mathbf{S}_K$ via ℓ_2 norm. We continue computing the correlation between each dictionary atom $\boldsymbol{a}_i$ and the nonlocal joint hyperspectral signal matrix $\mathbf{S}_K$ of the current test pixel $\boldsymbol{s}_c$

$$\text{corr}_i = \|\boldsymbol{a}_i' \bullet \mathbf{S}_K\|_1 \tag{15}$$

where $\boldsymbol{a}_i'$ is the transpose of $\boldsymbol{a}_i$, which is the ith atom of the whole dictionary, and corr_i is the corresponding correlation value. We acquire a correlation set $corr\_set$, in which each element reflects the correlation between the dictionary atom and the current test pixel. We then sort $corr\_set$ in a descending order and select the first L atoms from the global dictionary, which are considered as the active ones and are used to construct the locally adaptive dictionary $\mathbf{A}^L$ for the test pixel $\boldsymbol{s}_c$. In addition, we also set an indicator set I with N elements, where $I_i = 1$ means that the ith atom of $\mathbf{A}$ is active and can be found in $\mathbf{A}^L$. Thus, it has the remarkable property that, under very mild conditions, the error rate of the K-NN selection method tends to the Bayes optimal value as the sample size tends to infinity [36].

In the hyperspectral classification case, the following aspects should be considered. First, the locally adaptive subdictionary $\mathbf{A}^L$ for each test pixel is specific, so we should preconstruct it before the joint collaborative coding of the test pixel. The second tip is that the specific new subdictionary does not need to be overcomplete and is much smaller than the original dictionary $\mathbf{A}$, so the linear CR process of the test pixel is quicker and more stable than the JSRC method [14].

D. Final Classification Scheme of HSI

To summarize these three sections, the NJCRC-LAD can be represented as

$$\hat{\boldsymbol{\Psi}}^L = \arg\min_{\Psi^l} \left\{ \|\mathbf{S}_K - \mathbf{A}^L \boldsymbol{\Psi}^L\|_F^2 + \lambda\|\boldsymbol{\Psi}^L\|_F^2 \right\} \quad (16)$$

where $\hat{\boldsymbol{\Psi}}^L$ refers to the local coefficient matrix corresponding to the locally adaptive dictionary $\mathbf{A}^L$. The coefficient matrix $\hat{\boldsymbol{\Psi}}^L$ can be solved via (12).

We next acquire the global coefficient matrix $\hat{\boldsymbol{\Psi}}^g \in \mathbb{R}^{N \times K}$ and $N = \sum_i^M N_i$ by copying the row vectors corresponding to the active dictionary atoms from $\boldsymbol{\Psi}^L$ and setting the other rows as zero. The specific scheme for the global coefficient matrix construction is shown in Algorithm 1.

Once the global coefficient matrix is obtained, the label of the center pixel $\boldsymbol{s}_c$ is determined by the following classification rule:

$$class(\boldsymbol{s}_c) = \arg\min_{i=1,\ldots,M} \left\{ \|\mathbf{S}_K - \mathbf{A}_i \hat{\boldsymbol{\Psi}}_i^g\|_F / \|\hat{\boldsymbol{\Psi}}_i^g\|_F \right\} \quad (17)$$

where $\mathbf{A}_i$ is a subpart of $\mathbf{A}$ associated with class i and $\hat{\boldsymbol{\Psi}}_i^g$ denotes the portion of the recovered collaborative coefficients $\hat{\boldsymbol{\Psi}}^g$ for the ith class.

Algorithm 1: Global coefficient matrix $\hat{\boldsymbol{\Psi}}^g$ construction

---

Input: 1) The local coefficient matrix $\hat{\boldsymbol{\Psi}}^L \in \mathbb{R}^{L \times K}$; and 2) indicator set I with N elements, and $I_i = 0$ or 1, for $i = 1,\ldots,N$, in which "1" means that the corresponding dictionary atom is active and "0" means inactive.
Initialization: Set the initial global coefficient matrix $\hat{\boldsymbol{\Psi}}^g \in \mathbb{R}^{N \times K}$ as a zero matrix, and set an indicator $v = 1$
    For $i = 1$ **to** N
      if $I_i = 1$;
        $\hat{\boldsymbol{\Psi}}^g(i,:) = \hat{\boldsymbol{\Psi}}^L(v,:)$;
        $v{+}{+}$;
      End if
    End For
Output: The global coefficient matrix $\hat{\boldsymbol{\Psi}}^g$

The implementation details of the proposed NJCRC-LAD algorithm are shown in Algorithm 2.

Algorithm 2: The NJCRC-LAD algorithm for HSI classification

Input: 1) An HSI containing training samples, in which the test pixel located at p can be represented as $\boldsymbol{s}_p \in \mathbb{R}^B$, where B denotes the number of bands
    2) Parameters: Column number L of the compact subdictionary $\mathbf{A}^L$, regulation parameter λ, spatial neighborhood size T, and the number K of spectral signals in the joint-signal matrix
Initialization: Construct the entire dictionary $\mathbf{A}$ by stacking all the training samples in this HSI, and normalize every column of $\mathbf{A}$ to have a unit ℓ_2 norm

For each pixel in the HSI:
    1) Construct the initial joint-signal matrix $\mathbf{S} = [\boldsymbol{s}_1 \ \boldsymbol{s}_2 \ \cdots \ \boldsymbol{s}_T] \in \mathbb{R}^{B \times T}$, where p is located at the center of the neighborhood window, and normalize the columns of $\mathbf{S}$ to have a unit ℓ_2 norm
    2) Select and construct the nonlocal joint collaborative matrix $\mathbf{S}_K$ among $\mathbf{S}$ via the K-NN approach
    3) Select and stack the L atoms in $\mathbf{A}$ over the indicator set I to construct the locally adaptive compact dictionary $\mathbf{A}^L$
    4) Code $\mathbf{S}_K$ over $\mathbf{A}^L$ with (16)
    5) Construct the global coefficient matrix $\hat{\boldsymbol{\Psi}}^g$ via Algorithm 1
    6) Compute the residuals and label the test pixel with (17)
    7) Turn to the next test pixel
End For
Output: A 2-D matrix which records the labels of the HSI

IV. EXPERIMENTS

In this section, we investigate the effectiveness of the proposed NJCRC-LAD algorithm with three HSIs. The classifiers of SVM with radial basis function kernel [2], [5], [36], CRC [21], SRC with an improved ℓ_1-norm algorithm called the least absolute shrinkage and selection operator (LASSO) [37], [38], and JSRC with a greedy pursuit algorithm (referred to as simultaneous orthogonal matching pursuit in [14]) are used as benchmarks in this paper. In addition, two simplified versions of the proposed NJCRC-LAD method are also included in the comparisons. One is the CR classification with a locally adaptive dictionary (referred to as CRC-LAD), which can be considered as a special case of NJCRC-LAD where the nonlocal joint-signal matrix is just the central test pixel itself. The other one is the nonlocal joint CR classification algorithm, which uses the general dictionary in the joint CR signal reconstruction process. For simplicity, the SRC and JSRC methods are referred to as sparsity-based methods, and the CRC, CRC-LAD, NJCRC, and NJCRC-LAD algorithms are called collaboration-based methods. All the experiments were carried out using MATLAB on a computer machine with one 2.20-GHz processor and 4.0 GB of RAM.

A. AVIRIS Data Set: Indian Pines Image

This scene was gathered by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over the Indian Pines test site in Northwest Indiana and consists of 145×145 pixels and 220 spectral reflectance bands in the wavelength range 0.4–2.5 μm. The false-color composite of the Indian Pines image is shown in Fig. 1(a). We also reduced the number of bands to 200 by removing bands covering the regions of water absorption: 104–108, 150–163, and 220, as referred to in [34]. In this experiment, we randomly sampled 9% of the data in each class as the training samples and the remainder as the test samples. This image contains 16 ground-truth classes, and the numbers of the training and test sets are shown in Table I. The training and test sets are visually shown in Fig. 1(b) and (c), respectively.
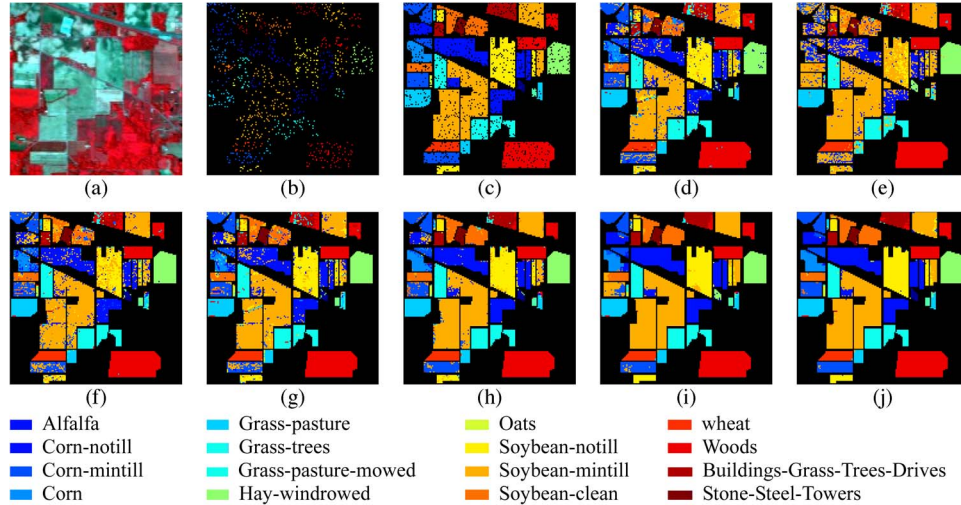
Fig. 1.   Classification results for the Indian Pines image: (a) False-color image (R: 57, G: 27, B: 17), (b) training set, (c) test set, (d) SVM, (e) CRC, (f) SRC, (g) CRC-LAD, (h) JSRC, (i) NJCRC, and (j) NJCRC-LAD.

TABLE I
SIXTEEN GROUND-TRUTH CLASSES OF THE AVIRIS INDIAN PINES
DATA SET AND THE TRAINING AND TEST SETS FOR EACH CLASS

No.	Class name	Samples	
		Train	Test
1	Alfalfa	6	40
2	Corn-notill	129	1299
3	Corn-mintill	83	747
4	Corn	24	213
5	Grass-pasture	48	435
6	Grass-trees	73	657
7	Grass-pasture-mowed	5	23
8	Hay-windrowed	48	430
9	Oats	4	16
10	Soybean-notill	97	875
11	Soybean-mintill	196	2259
12	Soybean-clean	59	534
13	Wheat	21	184
14	Woods	114	1151
15	Bldgs-grass-trees-drives	39	347
16	Stone-steel-towers	12	81
	Total	958	9291

The regularized parameter λ for these four ℓ_2-norm-based classification algorithms (we consider that the Frobenius norm is an ℓ_2 norm for the matrix in this paper) ranges from 1×10^{-7} to 1×10^{-2}. The column number L of the locally adaptive subdictionary $\mathbf{A}^L$ in CRC-LAD and NJCRC-LAD ranges from $L = 10$ to $L = 120$. For the NJCRC and NJCRC-LAD algorithms, the number of joint signals K is chosen between $K = 5$ and $K = 60$, and the neighborhood window size T ranges from 9 to 289. The optimal parameter setting for the NJCRC-LAD method is as follows: $L = 110$, $\lambda = 1 \times 10^{-5}$, $T = 81$, and $K = 45$. The corresponding optimal parameter setting for CRC-LAD is $\lambda = 1 \times 10^{-5}$ and $L = 55$; for NJCRC, it is $\lambda = 1 \times 10^{-5}$, $T = 81$, and $K = 45$; and for CRC, it is $\lambda = 1 \times 10^{-5}$. The parameters for SVM and SRC are set as the corresponding optimal parameters, and the optimal size of the spatial neighborhood in JSRC is 25.

The classification maps of the various classification methods are visually shown in Fig. 1(d)–(j). The quantitative accuracy results, which consist of the classification accuracy for every class, the overall accuracy, and the kappa coefficient,

are shown in Table II. Comparing the classification results of NJCRC over those of CRC, it can be clearly observed that the nonlocal neighborhood information can indeed support the classification performance. The improvement of CRC-LAD over CRC suggests that the locally adaptive dictionary is also helpful in improving the classification accuracy. It can also be observed that, by simultaneously incorporating the nonlocal neighborhood information and utilizing the locally adaptive subdictionary, the proposed NJCRC-LAD algorithm leads to the best classification map among all the classifiers.

In this image, we consider Classes 1, 4, 7, 9, 15, and 16 as small classes, and the size of each class is shown in Table I. It can be seen that the CRC classification results for these classes are quite poor, as shown in Table II, and are even zero in Classes 1, 7, and 9. It seems that CR cannot work well for these classes. In fact, this phenomenon is mainly caused by the lack of training samples for the associated class. In this case, the ℓ_2-norm regularization will magnify the weight of the interference atoms and will finally misidentify the pixel, while SRC with the sparsity norm can alleviate this problem by restricting the support to a subset of the dictionary, not the entire dictionary. It can also be observed that this problem is alleviated with the help of the locally adaptive dictionary in the CRC-LAD and NJCRC-LAD methods. In addition, the locally adaptive dictionary imposes the locality constraint, which is equivalent or even superior to the "sparsity constraint" in the sparsity-based algorithms [23].

We next demonstrate the relationship between CR, sparsity, and the locally adaptive dictionary constraint in Figs. 2 and 3. We randomly select a test pixel which belongs to Class 6 and is located at (111, 75) in the Indian Pines HSI and represent it by SRC with ℓ_0- and ℓ_1-norm regularizations, CRC with ℓ_2-norm regularization, and CRC-LAD with ℓ_2-norm regularization. We also calculate the coefficients of the patch pixels around the test one using the joint sparse representation algorithm with $\ell_{\text{row},0}$- and $\ell_{1,2}$-norm regularizations, NJCRC with Frobenius-norm regularization, and NJCRC-LAD with Frobenius-norm regularization. Fig. 2 shows the recovered coefficients of the current test pixel under the various norms, and Fig. 3 shows

TABLE II
CLASSIFICATION ACCURACY FOR THE INDIAN PINES IMAGE WITH THE TEST SET

| Class	SVM	CRC	SRC	CRC -LAD	JSRC	NJCRC	NJCRC -LAD
1	72.50	0	82.50	82.50	92.50	30.00	**97.50**
2	76.98	64.82	79.37	77.89	91.45	**96.07**	95.61
3	75.77	14.73	74.43	72.56	85.27	76.04	**97.86**
4	60.56	00.47	72.30	57.75	81.22	71.36	**93.43**
5	90.80	64.37	87.59	84.37	91.72	91.26	**94.48**
6	96.96	94.37	98.02	97.56	**100**	**100**	99.70
7	91.30	0	73.91	34.78	78.26	**100**	**100**
8	96.05	99.53	99.77	**100**	**100**	**100**	**100**
9	68.75	0	75.00	81.25	31.25	43.75	87.50
10	76.91	21.83	72.34	81.94	87.66	94.17	**97.60**
11	81.45	94.33	85.88	87.34	94.64	**99.34**	96.72
12	64.23	21.72	75.84	74.53	91.95	88.39	**97.57**
13	91.30	96.74	**100**	99.46	**100**	98.91	98.91
14	95.48	99.30	98.35	98.44	98.78	**100**	99.74
15	53.31	29.97	61.96	56.48	89.63	**94.52**	89.05
16	92.59	87.65	98.77	96.30	**100**	97.53	98.77
OA	81.63	66.89	84.45	84.47	93.13	94.44	**97.12**
Kappa	79.01	60.37	82.18	82.17	92.15	93.60	**96.71**
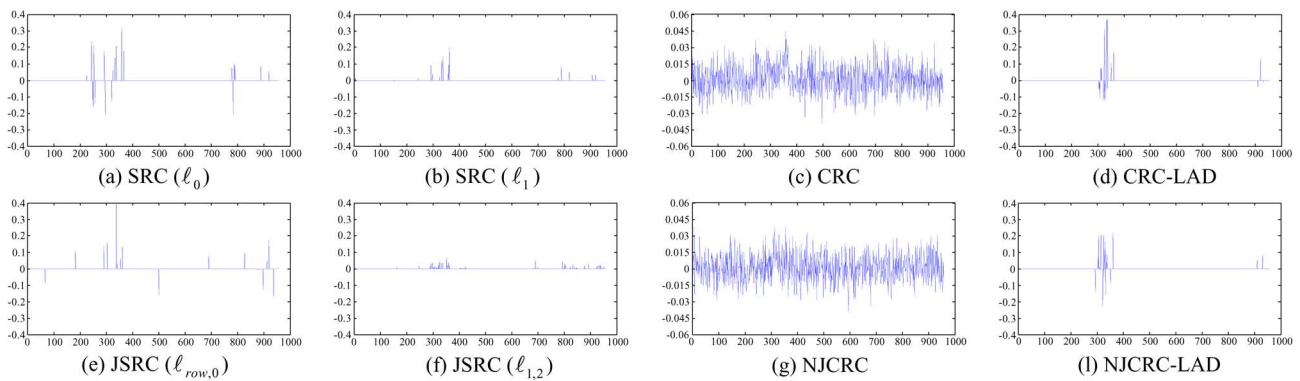
Fig. 2. Estimated construction coefficients for the pixel located at (111, 75) in the Indian Pines image. (a) Estimated sparse ℓ_0 coefficients in SRC (the orthogonal matching pursuit [39] algorithm), (b) estimated sparse ℓ_1 coefficients in SRC (the LASSO algorithm), (c) estimated ℓ_2 coefficients in the CRC algorithm, (d) estimated locality-constrained ℓ_2 coefficients in the CRC-LAD algorithm, (e) estimated $\ell_{row,0}$-norm coefficients with the spatial JSM (referred to as SOMP), (f) estimated $\ell_{1,2}$-norm coefficients with spatial joint sparse representation (the L1/L2-regularization based on block-coordinate descent algorithm can also be found in [40]), (g) estimated Frobenius-norm-based coefficients with the nonlocal joint collaboration model (referred to as the NJCRC algorithm), and (l) estimated locality-constrained Frobenius-norm-based coefficients with the nonlocal joint collaboration model (referred to as the NJCRC-LAD algorithm). The MATLAB codes of all the sparsity-based algorithms that we use here can be freely downloaded from [40].
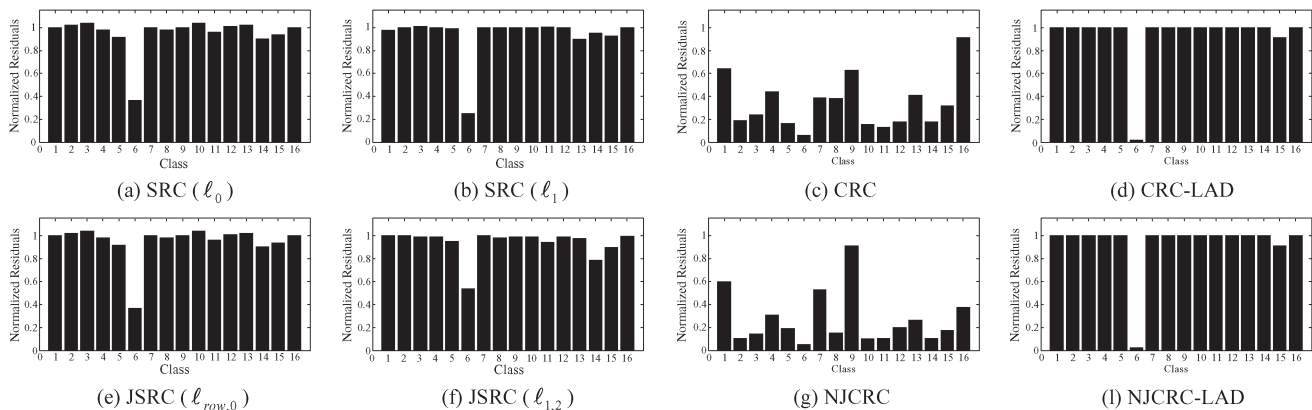

Fig. 3. Normalized residuals for each class for the pixel located at (111, 75) in the Indian Pines image. (a) SRC algorithm with the ℓ_0-norm constraint, (b) SRC with the ℓ_1-norm constraint, (c) CRC with the ℓ_2-norm constraint, (d) CRC-LAD with the ℓ_2-norm regularization and locally adaptive dictionary constraint, (e) JSRC algorithm with the $\ell_{row,0}$-norm constraint, (f) estimated $\ell_{1,2}$-norm coefficients with spatial joint sparse representation, (g) NJCRC with the Frobenius-norm constraint, and (l) NJCRC-LAD with the Frobenius-norm regularization and locally adaptive dictionary constraint.

their corresponding residuals. It can be observed in Fig. 3 that all the classifiers can identify the pixel properly, but the coefficient vectors for each algorithm are largely different, as shown in Fig. 2. For the ℓ_0 norm in SRC and its matrix form, the $\ell_{row,0}$ norm in JSRC, it can be seen that the nonzero coef-

ficients mainly locate in Classes 6 and 13–15, and the residuals associated with Class 6 have the smallest value, which suggests that the dominant contributions mainly come from the training samples associated with Class 6. For the ℓ_1 norm in SRC and its matrix form, the $\ell_{1,2}$ norm, the sparse vector illustrated in
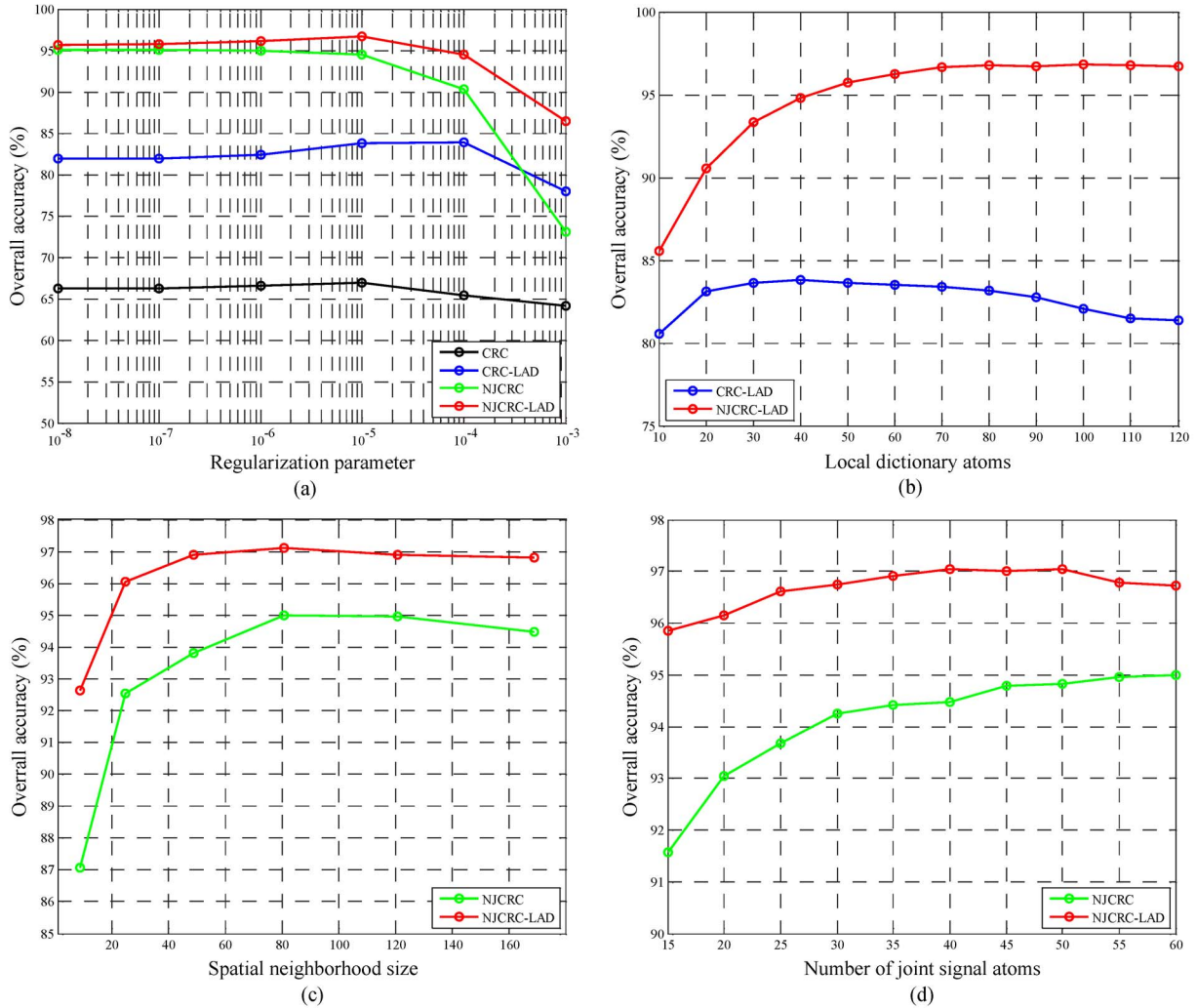
Fig. 4.   Classification results versus parameters for the four collaboration-based algorithms. (a) Regularization parameter λ for CRC, CRC-LAD, NJCRC, and NJCRC-LAD. (b) Number of locally adaptive dictionary atoms L for CRC-LAD and NJCRC-LAD. (c) Size of the spatial neighborhood T for NJCRC and NJCRC-LAD. (d) Number of joint signals K in NJCRC and NJCRC-LAD.

<div style="display:flex">

<div>

TABLE III
SPEED WITH THE INDIAN PINES HYPERSPECTRAL IMAGE
FOR THE SINGLE-SIGNAL ALGORITHMS

	CRC	SRC	CRC-LAD
Times (s)	9.5530	64.4765	13.4225
Speed-up	$6.7493 = \dfrac{64.4765}{9.5530}$		$4.8036 = \dfrac{64.4765}{13.4225}$

</div>

<div>

TABLE IV
SPEED WITH THE INDIAN PINES HYPERSPECTRAL IMAGE
FOR THE JOINT-SIGNAL ALGORITHMS

Neighborhood size	JSRC	NJCRC-LAD	Speed-up
9	79.4763	29.8539	2.6628
25	**154.5468**	39.9083	3.8723
49	279.5702	52.4832	5.3269
81	441.8671	**66.8165**	6.6131
121	646.9860	84.0632	7.6964
169	941.3547	109.5890	8.5899
Optimal speed-up	$2.3129 = \dfrac{154.5468}{66.8165}$		

</div>

</div>

Fig. 2(b) and (f) shows a similar observation to the ℓ_0-norm case. We can also see that the training samples in Class 6 contribute the most and the training samples in Classes 13–15 also work as collaborative partners. Although the CRC labeling is correct for this pixel, as shown in Fig. 3(c), the difference between the residuals of all the classes is not large, which is the reason why CRC achieves the worst classification result, as shown in Table II. With the locally adaptive dictionary, the CRC-LAD and NJCRC-LAD algorithms achieve sparse coefficients and the best residual distributions. It should be noted that all the coefficients shown in Fig. 2 have nonzero values associated with other classes, except for Class 6. That is to say, the training pixels from the other classes also participate in the lin-

ear representation of the test pixel, which is the working mechanism of the CR. We also compare Fig. 3(d) with Fig. 3(c), which suggest that the dictionary is pruned by K-NN in the locally adaptive dictionary construction.

We next investigate the effects of the regularization parameter λ for the four collaboration-based classifications (CRC, CRC-LAD, NJCRC, and NJCRC-LAD). For NJCRC-LAD, we fix the rest of the parameters as $K = 45$, $T = 81$, and
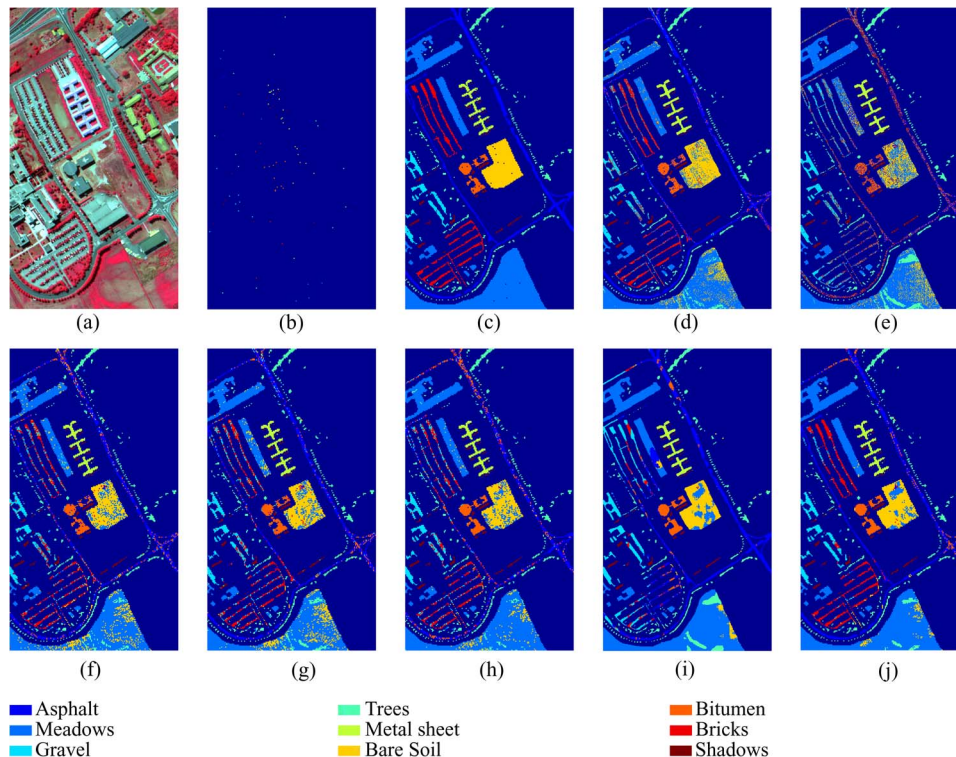
Fig. 5.   Classification results with the University of Pavia image: (a) False-color image (R: 102, G: 56, B: 31), (b) training set, (c) test set, (d) SVM, (e) CRC, (f) SRC, (g) CRC-LAD, (h) JSRC, (i) NJCRC, and (j) NJCRC-LAD.

$L = 110$. In Fig. 4(a), the horizontal axis indicates the regularization parameter λ, and the vertical axis shows the corresponding optimal overall accuracy (in percent) of the different collaboration-based algorithms. It can be clearly observed that the proposed NJCRC-LAD method yields the best accuracy. With the increase in λ, the performances of the algorithms improve quite slowly and then quickly decline when λ exceeds a certain threshold, which suggests that the accuracies for these collaboration-based algorithms are robust with regard to the regularization parameter λ.

We then focus on the effects of the locally adaptive dictionary for the CRC-LAD and NJCRC-LAD algorithms. In Fig. 4(b), we fix the rest of the parameters for NJCRC-LAD as $\lambda = 1 \times 10^{-5}, T = 81$, and $K = 45$ and set $\lambda = 1 \times 10^{-5}$ for CRC-LAD. It can be observed that both the plots rise quickly and reach a maximum point and then become relatively stable, and the performance of NJCRC-LAD is better and more stable than that of CRC-LAD.

The joint spatial aspects of the NJCRC-LAD algorithm contain two parameters: the neighborhood size T and the number of nonlocal hyperspectral signals K. We fix $\lambda = 1 \times 10^{-5}$ and $L = 110$ and show the plots of the classification results versus the two parameters in Fig. 4(c) and (d), respectively. Both plots rise quickly and reach a maximum point and then remain relatively stable with only a tiny decline, which shows the robustness of the proposed NJCRC-LAD method.

We next compare the running times of the various classification algorithms. Table III shows the running times of the three single-signal algorithms (CRC, CRC-LAD, and SRC with a fast ℓ_1-minimization method named LASSO), and Table IV shows

TABLE V
NINE GROUND-TRUTH CLASSES IN THE ROSIS UNIVERSITY OF PAVIA DATA SET AND THE TRAINING AND TEST SAMPLE SETS FOR EACH CLASS

No.	Class name	Samples	
		Train	Test
1	Asphalt	50	6581
2	Meadows	50	18599
3	Gravel	50	2049
4	Trees	50	3014
5	Metal sheet	50	1295
6	Bare Soil	50	4979
7	Bitumen	50	1280
8	Brick	50	3632
9	Shadows	50	897
	Total	450	42326

the running times of the two joint-signal algorithms (JSRC and NJCRC-LAD). The optimal classification accuracies can be found in Table II. In Tables III and IV, the speedup denotes the ratio of the processing time of the CR-related algorithm to that of the corresponding SR-related approach under the same neighborhood size, and the optimal speedup denotes the ratio of the processing time of the CR-related algorithm to that of the SR-related method under the corresponding optimal neighborhood size. For the Indian Pines image, CRC is the fastest but gives the worst classification performance, while CRC-LAD achieves comparable classification accuracy and is faster than SRC. The bold font in Table IV denotes the running time that each algorithm requires when reaching its optimal classification performance. In Table IV, it can be observed that the running time of the proposed NJCRC-LAD method is less

TABLE VI
CLASSIFICATION ACCURACY (IN PERCENT) FOR THE UNIVERSITY OF PAVIA IMAGE WITH THE TEST SET

Class	SVM	CRC	SRC	CRC-LAD	JSRC	NJCRC	NJCRC-LAD
1	**85.88**	38.75	71.60	77.68	61.51	81.25	84.62
2	81.42	79.09	85.76	83.79	**95.56**	88.38	94.75
3	78.62	82.19	80.14	84.14	88.04	93.51	**94.83**
4	85.67	93.60	91.37	95.52	92.17	90.48	**97.35**
5	74.75	99.85	99.92	99.92	**100**	98.84	**100**
6	**81.30**	51.20	68.75	73.39	69.85	74.53	79.57
7	8625	65.70	85.47	87.81	94.37	96.25	**97.34**
8	76.18	11.45	60.93	63.77	55.07	14.95	**87.47**
9	**100**	93.98	97.21	92.75	96.99	98.89	95.76
OA	82.16	65.46	80.22	81.56	83.29	80.52	**91.21**
Kappa	76.92	55.61	74.09	75.99	77.74	74.35	**88.35**

than that of the JSRC algorithm with the same neighborhood size, and the speedup becomes greater with the increase in the neighborhood size. The optimal size for JSRC is smaller than that for NJCRC-LAD, which indicates the effectiveness of the nonlocal signal selection approach.

B. ROSIS Urban Data: University of Pavia, Italy

This scene was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor during a flight campaign over Pavia, Northern Italy. The number of spectral bands is 103, and the image size is 610×610 pixels. We cut a patch sized 610×340 pixels from the original image. The geometric resolution is 1.3 m. The false-color composite of the University of Pavia image is shown in Fig. 5(a). This image contains nine ground-truth classes, as shown in Table V. We randomly sampled 50 pixels in each class as the training samples and the remainder as the test samples. The training and test sets can be seen visually in Fig. 5(b) and (c), respectively.

The classification results using SVM, CRC, SRC, CRC-LAD, JSRC, NJCRC, and NJCRC-LAD are visually shown in Fig. 5(d)–(j), respectively. The quantitative evaluations, which consist of the classification accuracy for each class, the overall accuracy, and the kappa coefficient, are shown in Table VI. The optimal parameters for the NJCRC-LAD method are as follows: $K = 35$, $L = 100$, $\lambda = 1 \times 10^{-3}$, and $T = 81$. The corresponding optimal neighborhood size for JSRC is $T = 25$. From Table VI, it can be clearly observed that the NJCRC and CRC-LAD methods achieve large improvements over the original CRC method, which confirms the effectiveness of the proposed nonlocal signal matrix construction method and the locally adaptive dictionary atom selection strategy, respectively. By integrating the two techniques, the proposed NJCRC-LAD method yields the best overall accuracy, kappa coefficient, and classification accuracy for the most classes among the various state-of-the-art classification methods.

Tables VII and VIII show the running times with the University of Pavia image for the single- and joint-signal algorithms, respectively. Generally speaking, the running times for the collaboration-based algorithms are faster than those of the sparsity-based algorithms. The bold font in Table VIII denotes the running time that each algorithm takes when reaching its optimal classification performance. The reason that the optimal speedup is less than one is mainly caused by the different spatial neighborhood sizes for these two algorithms. In addition, in the

TABLE VII
SPEED WITH THE UNIVERSITY OF PAVIA IMAGE
FOR THE SINGLE-SIGNAL ALGORITHMS

	CRC	SRC	CRC-LAD
Times (s)	9.5530	64.4765	13.4225
Speed-up	$4.1448 = \dfrac{61.4054}{14.8151}$		$1.3871 = \dfrac{61.4054}{44.2705}$

TABLE VIII
SPEED WITH THE UNIVERSITY OF PAVIA IMAGE
FOR THE JOINT-SIGNAL ALGORITHMS

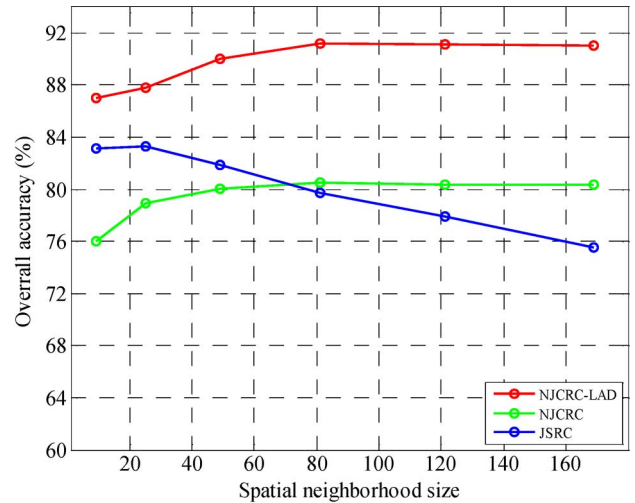| Neighborhood size	JSRC	NJCRC-LAD	Speed-up
9	80.4741	83.9359	0.959
25	**150.8502**	120.8422	1.248
49	239.9291	159.8595	1.501
81	373.4729	**191.2628**	1.953
121	531.9892	251.4817	2.115
169	742.1156	309.7660	2.396
Optimal speed-up	$0.7887 = \dfrac{150.8502}{191.2628}$		

Fig. 6. Classification results versus spatial neighborhood size for the three joint spatial information algorithms.

proposed NJCRC-LAD method, the selection method for both the nonlocal hyperspectral signals and locally adaptive dictionary atoms is the K-NN method, which is quite time consuming and occupies most of the running time of the proposed NJCRC-LAD classification method.
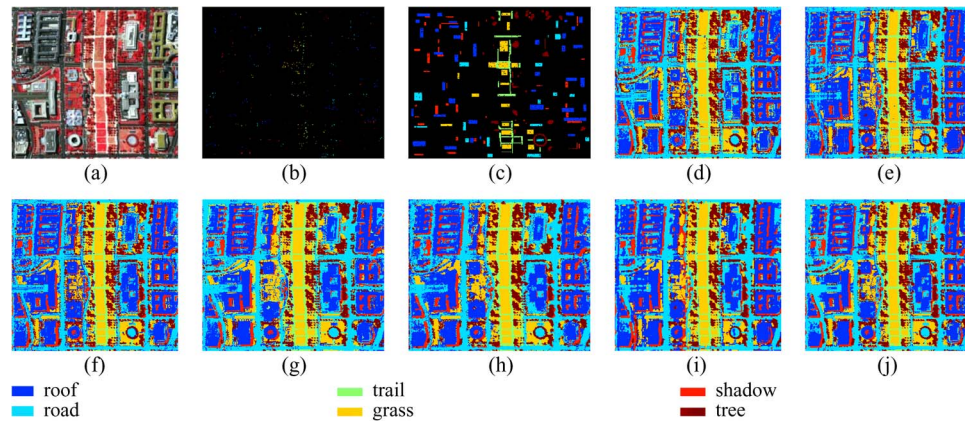
Fig. 7. Classification results with the Washington DC image: (a) False-color image (R: 63, G: 52, B: 36), (b) training set, (c) test set, (d) SVM, (e) CRC, (f) SRC, (g) CRC-LAD, (h) JSRC, (i) NJCRC, and (j) NJCRC-LAD.

The plot in Fig. 6 shows the classification overall accuracy versus the neighborhood size for the three joint spatial information algorithms. The horizontal axis indicates the spatial neighborhood size T, and the vertical axis shows the corresponding optimal overall accuracy (in percent) of the different joint algorithms. It can be seen that the proposed NJCRC-LAD method yields the best overall accuracy for all the neighborhood sizes and shows the best robustness.

C. HYDICE Data Set: Washington DC Image

This image is a part of an airborne hyperspectral data flight line over the Washington DC Mall, which was acquired by the Hyperspectral Digital Image Collection Experiment (HYDICE) sensor and is provided with the permission of the Spectral Information Technology Application Center of Virginia, which was responsible for its collection. The sensor system used in this case measured a pixel response in 210 bands in the 0.4- to 2.4-μm region of the visible and infrared spectrum. Bands in the 0.9- and 1.4-μm regions where the atmosphere is opaque have been omitted from the data set, leaving 191 bands. The data set contains 280 scan lines, with 307 pixels in each scan line. The false-color composite of the Washington DC image is shown in Fig. 7(a). This image contains six ground-truth classes, as shown in Table IX. For each of the six classes, we randomly chose around 5% of the labeled samples for training and the rest for testing. The training and test sets are visually shown in Fig. 7(b) and (c), respectively. In addition, the numbers of training and test samples are also shown in Table IX.

The classification quantitative evaluation results are summarized in Table X, and the classification maps are visually shown in Fig. 7(d)–(j). The optimal parameters for NJCRC-LAD are as follows: $K = 9$, $L = 110$, $\lambda = 1 \times 10^{-4}$, and $T = 49$. The corresponding optimal neighborhood size for JSRC is $T = 9$. It can be seen that the proposed NJCRC-LAD method achieves the best overall classification result. The running times for all the methods are shown in Tables XI and XII, which show similar observations to the previous two experiments. Fig. 8 shows the classification overall accuracies versus the neighborhood size for the three joint spatial information algorithms.

TABLE IX
SIX GROUND-TRUTH CLASSES OF THE HYDICE WASHINGTON DC
MALL DATA SET AND THE TRAINING AND TEST SETS FOR EACH CLASS

No.	Class name	Samples	
		Train	Test
1	Roof	156	2973
2	Road	70	1332
3	Trail	63	1201
4	Grass	90	1700
5	Shadow	56	1064
6	Tree	60	1134
	Total	495	9404

It is demonstrated that the classifiers with nonlocal signal matrix selection tend to be more robust than JSRC with the increase in the neighborhood size. It is also concluded that the classification performance for NJCRC-LAD is the best among all the classifiers, with excellent robustness.

V. CONCLUSION

In this paper, we have proposed a novel NJCRC-LAD for HSI classification. First, we utilize the CR mechanism and build a JCM to incorporate neighborhood pixels under the joint collaboration assumption. Next, to reject the neighboring pixels that are dissimilar to the central test one, a nonlocal joint-signal selection method is introduced to better fulfill the joint collaboration assumption. Finally, a subdictionary is constructed, which is adaptive to the nonlocal signal matrix, to replace the general dictionary. The proposed NJCRC-LAD method was tested on three HSIs. The extensive experimental results clearly show that the proposed NJCRC-LAD method achieves superior classification performance.

However, the proposed algorithm still has room for improvement. For instance, the method used to determine the nonlocal joint signals and the active locally adaptive atoms is the K-NN method, which is simple but time consuming. K-NN increases the computational burden, particularly in the high-spatial-resolution case. Other smart methods for the selection with low complexity could be taken into consideration.

TABLE X
CLASSIFICATION ACCURACY (IN PERCENT) FOR THE WASHINGTON DC IMAGE WITH THE TEST SET

Class	SVM	CRC	SRC	CRC-LAD	JSRC	NJCRC	NJCRC-LAD
1	87.39	97.28	98.45	97.34	98.39	97.78	**98.52**
2	90.32	89.71	93.09	97.60	99.62	90.62	**99.77**
3	92.01	94.34	**98.83**	96.59	98.42	96.34	98.33
4	97.76	98.88	98.88	99.53	99.41	99.65	**99.94**
5	99.44	96.99	**99.81**	99.15	87.41	97.84	99.09
6	93.30	97.18	96.38	97.44	**98.41**	96.65	97.88
OA	92.37	96.08	97.72	97.89	97.51	96.79	**98.92**
Kappa	90.54	95.10	97.16	97.38	96.90	96.00	**98.65**

TABLE XI
SPEED WITH THE WASHINGTON DC IMAGE
FOR THE SINGLE-SIGNAL ALGORITHMS

	CRC	SRC	CRC-LAD
Times (s)	4.8283	59.6503	13.6186
Speed-up	$12.3543 = \frac{59.6503}{4.8283}$		$4.3800 = \frac{59.6503}{13.6186}$

TABLE XII
SPEED WITH THE WASHINGTON DC IMAGE
FOR THE JOINT-SIGNAL ALGORITHMS

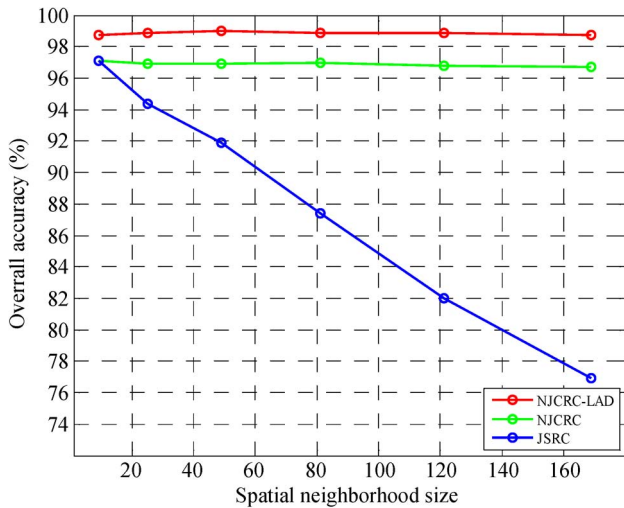| Neighborhood size	JSRC	NJCRC-LAD	Speed-up
9	**28.8663**	21.5688	1.3383
25	51.8743	24.7132	2.0991
49	87.5973	**39.3947**	2.2236
81	102.4696	46.1742	2.2192
121	149.0785	71.0215	2.0991
169	229.7543	96.4305	2.2893
Optimal speed-up	$0.7327 = \frac{28.8663}{39.3947}$		

Fig. 8.   Classification results versus spatial neighborhood size for the three joint spatial information algorithms.

ACKNOWLEDGMENT

The authors would like to thank Prof. D. Landgrebe from Purdue University for providing the AVIRIS image of Indian Pines and the Hyperspectral Digital Image Collection Experiment image of Washington DC Mall, Prof. Gamba from the University of Pavia for providing the Reflective Optics System Imaging Spectrometer data set, and Dr. J. Mairal from the Institut National de Recherche en Informatique et en Automatique for sharing the SPArse Modeling Software package. The authors would also like to thank the handling editor and anonymous reviewers for their careful reading and helpful remarks.

REFERENCES

[1] J. C. Harsanyi and C. I. Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 4, pp. 779–785, Jul. 1994.
[2] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gambah, A. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, no. 1, pp. S110–S122, Sep. 2009.
[3] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral–spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 809–823, Mar. 2012.
[4] Y. Tarabalka, J. A. Benediktsson, J. Chanussot *et al.*, "Multiple spectral–spatial classification approach for hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4122–4132, Nov. 2010.
[5] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.
[6] B. Demir and S. Erturk, "Hyperspectral image classification using relevance vector machines," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 4, pp. 586–590, Oct. 2007.
[7] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.
[8] J. Wright, A. Y. Yang, A. Ganesh *et al.*, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.
[9] C. Lang, G. Liu, J. Yu, and S. Yan, "Saliency detection by multitask sparsity pursuit," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1327–1338, Mar. 2012.
[10] T. Guha and R. Ward, "Learning sparse representations for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 8, pp. 1576–1588, Aug. 2012.
[11] X. Mei and H. Ling, "Robust visual tracking and vehicle classification via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 11, pp. 2259–2272, Nov. 2011.
[12] W. Deng, J. Hu, and J. Guo, "Extended SRC: Undersampled face recognition via intraclass variant dictionary," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 9, pp. 1864–1870, Sep. 2012.
[13] E. Elhamifar and R. Vidal, "Robust classification using structured sparse representation," in *Proc. IEEE CVPR*, 2011, pp. 1873–1879.
[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, 2011.
[15] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification via kernel sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 217–231, Jan. 2013.
[16] Q. Sami ul Haq, L. Tao, F. Sun, and S. Yang, "A fast and robust sparse approach for hyperspectral data classification using a few labeled samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 6, pp. 2287–2302, Jun. 2012.
[17] Y. Qian, M. Ye, and J. Zhou, "Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet

texture features," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2276–2291, Apr. 2012.

[18] D. Baron, M. F. Duarte, M. B. Wakin, S. Sarvotham, and R. G. Baraniuk, *Distributed Compressive Sensing*, 2009, arXiv preprint arXiv:0901.3403.

[19] A. Plaza, P. Martinez, J. Plaza *et al.*, "Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 466–479, Mar. 2005.

[20] J. A. Benediktsson and I. Kanellopoulos, "Classification of multisource and hyperspectral data based on decision fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1367–1377, May 1999.

[21] L. Zhang, M. Yang, X. Feng, Y. Ma, and D. Zhang, *Collaborative Representation Based Classification for Face Recognition*, 2012, arXiv preprint arXiv:1204.2358.

[22] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, Mar. 2006.

[23] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *Proc. IEEE Conf. CVPR*, 2010, pp. 3360–3367.

[24] K. Fukunaga and D. M. Hummels, "Bayes error estimation using Parzen and k-NN procedures," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, no. 5, pp. 634–643, Sep. 1987.

[25] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *J. Mach. Learn. Res.*, vol. 10, pp. 207–244, Dec. 2009.

[26] E. Blanzieri and F. Melgani, "Nearest neighbor classification of remote sensing images with the maximal margin principle," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 6, pp. 1804–1811, Jun. 2008.

[27] S. J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale l1-regularized least squares," *IEEE J. Sel. Top. Signal Process.*, vol. 1, no. 4, pp. 606–617, Dec. 2007.

[28] A. Y. Yang, Z. Zhou, A. Ganesh, S. Shankar Sastry, and Y. Ma, *Fast l1-Minimization Algorithms for Robust Face Recognition*, 2010, arXiv preprint arXiv:1007.3753.

[29] S. Kong and D. Wang, *Online Discriminative Dictionary Learning for Image Classification Based on Block-Coordinate Descent Method*, 2012, arXiv preprint arXiv:1203.0856.

[30] M. D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.

[31] P. Zhu, L. Zhang, Q. Hu, and S. C. K. Shiu, "Multi-scale patch based collaborative representation for face recognition with margin distribution optimization," in *Proc. Comput. Vis.–ECCV*, vol. 7572, *Lecture Notes in Computer Science*, A. Fitzgibbon, S. Lazebnik, P. Perona *et al.*, Eds., 2012, pp. 822–835, Springer-Verlag: Berlin, Germany.

[32] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2011, pp. 471–478.

[33] J. Waqas, Z. Yi, and L. Zhang, "Collaborative neighbor representation based classification using l2-minimization approach," *Pattern Recognit. Lett.*, vol. 34, no. 2, pp. 201–208, Jan. 2013.

[34] K. Yu, T. Zhang, and Y. Gong, "Nonlinear learning using local coordinate coding," *Adv. Neural Inf. Process. Syst.*, vol. 22, pp. 2223–2231, 2009.

[35] W. Dong, L. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 1838–1857, Jul. 2011.

[36] H. Zhang, A. C. Berg, M. Maire, and J. Malik, "SVM-KNN: Discriminative nearest neighbor classification for visual category recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2006, pp. 2126–2136.

[37] R. Tibshirani, "Regression shrinkage and selection via the LASSO: A retrospective," *J. R. Statist. Soc. B*, vol. 73, no. 3, pp. 273–282, Jun. 2011.

[38] T. Hastie, J. Taylor, R. Tibshirani, and G. Walther, "Forward stagewise regression and the monotone LASSO," *Electron. J. Statist.*, vol. 1, pp. 1–29, 2007.

[39] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. Conf. Rec. 27th Asilomar Conf. Signals, Syst. Comput.*, 1993, pp. 40–44.

[40] J. Mairal, F. Bach, J. Ponce, G. Sapiro, R. Jenatton, and G. Obozinski, SPAMS Software. [Online]. Available: http://www.di.ens.fr/willow/SPAMS/index.html

Jiayi Li (S'13) received the B.S. degree from Central South University, Changsha, China, in 2011. She is currently working toward the Ph.D. degree in the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China.

Her research interests include hyperspectral imagery, sparse representation, pattern recognition, and computation vision in remote sensing images.

Hongyan Zhang (M'13) 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.

Since 2010, he has been a Lecturer with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. His current research interests focus on image reconstruction and remote sensing image processing.

Yuancheng Huang received the B.S. degree from Chang'an University, Xi'an, China, in 2005 and the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China, in 2010.

He is currently a Lecturer with the College of Geomatics, Xi'an University of Science and Technology, Xi'an. His major research interests include pattern recognition, hyperspectral remote sensing, and image processing.

Liangpei Zhang (M'06–SM'08) received the B.S. degree in physics from Hunan Normal University, Changsha, China, in 1982, the M.S. degree in optics from the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China, in 1988, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1998.

He is currently with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, as the Head of the Remote Sensing Division. He is also a "Changjiang Scholar" Chair Professor appointed by the Ministry of Education of China. He is also currently a Principal Scientist for the China State Key Basic Research Project (2011–2016) appointed by the Ministry of Science and Technology of China to lead the remote sensing program in China. He also serves as an Associate Editor of the *International Journal of Ambient Computing and Intelligence*, the *International Journal of Image and Graphics*, the *International Journal of Digital Multimedia Broadcasting*, the *Journal of Geo-spatial Information Science*, and the *Journal of Remote Sensing (Chinese)*. He has more than 260 research papers and is the holder of five patents. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence.

Dr. Zhang is a Fellow of the Institution of Electrical Engineers, an Executive Member (Board of Governor) of the Chinese National Committee for the International Geosphere–Biosphere Programme, and an Executive Member for the China Society of Image and Graphics. He regularly serves as a Cochair of the series the International Society for Optical Engineering Conferences on Multispectral Image Processing and Pattern Recognition, Asian Conference on Remote Sensing, and many other conferences. He edits several conference proceedings, issues, and the Geoinformatics Symposiums. He also serves as an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.