

A Nonlocal Weighted Joint Sparse Representation Classification Method for Hyperspectral Imagery

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

Abstract—As a powerful and promising statistical signal modeling technique, sparse representation has been widely used in various image processing and analysis fields. For hyperspectral image classification, previous studies have shown the effectiveness of the sparsity-based classification methods. In this paper, we propose a nonlocal weighted joint sparse representation classification (NLW-JSRC) method to improve the hyperspectral image classification result. In the joint sparsity model (JSM), different weights are utilized for different neighboring pixels around the central test pixel. The weight of one specific neighboring pixel is determined by the structural similarity between the neighboring pixel and the central test pixel, which is referred to as a nonlocal weighting scheme. In this paper, the simultaneous orthogonal matching pursuit technique is used to solve the nonlocal weighted joint sparsity model (NLW-JSM). The proposed classification algorithm was tested on three hyperspectral images. The experimental results suggest that the proposed algorithm performs better than the other sparsity-based algorithms and the classical support vector machine hyperspectral classifier.

Index Terms—Classification, hyperspectral imagery, joint sparse representation, nonlocal weight.

I. INTRODUCTION

EVERY pixel in a hyperspectral image (HSI) is represented by hundreds of values, and each value corresponds to a different narrow wavelength [1]. These values constitute a high-dimensional vector with more dedicated spectral information. This property of HSI has opened up avenues for new remote sensing applications in various fields, such as precision agriculture [2], environmental protection [3], [4] and so on. Supervised HSI classification, which aims at categorizing the pixels in the image into one of several land-cover classes with representative training samples, is an important application. To date, a lot of different HSI classification techniques have been proposed, including independent component analysis (ICA) [4], decision trees (DT), and artificial neural networks (ANN) [2]. Among these approaches, the support vector machine (SVM) [5], [6],

which aims at discriminating two classes by fitting an optimal separating hyperplane to the training data within a multi-dimensional feature space, has shown excellent performance in supervised HSI classification. In recent years, many extended algorithms based on SVM have been proposed to improve the classification result. For example, Tsang *et al.* [7] presented a core vector machine (CVM) method which can optimize both the time and space complexities of the standard SVM algorithm for a large-scale data set; Bruzzone *et al.* [8] introduced a semi-supervised SVM-based classification method which exploits both labeled and unlabeled pixels; and Zhang *et al.* [9] proposed a hybrid of SVM and the KNN scheme.

In recent years, the sparsity of signals, which assumes that natural signals can be compactly represented by only a few coefficients that carry the most important information in a certain basis or dictionary, has become a powerful and promising statistical signal modeling tool for image processing and analysis. Sparse representation classification (SRC), first proposed by Wright *et al.* [10] for face recognition, has been widely used in various pattern recognition applications. Recently, Chen *et al.* [11] applied the SRC method to HSI classification and proposed the joint sparse representation classification (JSRC) method. This approach incorporates the spatial contextual information by the assumption of a joint sparsity model (JSM) [12], in that all the neighboring pixels around the central test pixel share a common sparsity pattern with different sets of coefficients. By incorporating the spatial neighborhood information in the classification process of the central test pixel, the JSRC method can achieve improved classification results. However, this method considers an equal contribution for any neighborhood pixel in the classification process of the central test pixel, which is more appropriate for pixels in a homogeneous area, and less reasonable for heterogeneous pixels, especially around image edges.

In view of this, in this paper, we propose a nonlocal weighted joint sparse representation classification (NLW-JSRC) method for hyperspectral imagery. The main contribution of this paper is that we consider different contributions for the neighboring pixels in the classification process of the central test pixel. That is, different weights are utilized for different neighboring pixels in the joint sparsity model (JSM). The weight of one specific neighboring pixel is determined by the structural similarity between the neighboring pixel and the central test pixel, which is referred to as a nonlocal weighting scheme, and has been widely used in image denoising applications [13]. With the self-similarity of the local structure [14] that the image itself offers, the nonlocal weighting scheme not only compares the spectrum at a single point but also the geometrical configuration of the whole neighborhood area [13]. After all the weights of the pixels in

Manuscript received April 03, 2013; revised May 02, 2013; accepted May 13, 2013. Date of publication June 03, 2013; date of current version August 01, 2014. This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2011CB707105, by the National Natural Science Foundation of China under Grants 61201342, 40930532 and 61102112. (Corresponding author: J. Li.)

H. Zhang, J. Li, and L. Zhang are with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China (e-mail: zhanghongyan@whu.edu.cn; zjjerica@163.com; zlp62@public.wh.hb.cn).

Y. Huang is with the College of Geomatics, Xi'an University of Science and Technology, Wuhan, China (e-mail: yuancheng.huang@gmail.com).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSTARS.2013.2264720

the neighborhood are determined, in this paper, we set up the nonlocal weighted joint sparsity model (NLW-JSM). The proposed NLW-JSRC method aims to ensure that the pixels which are similar to the central test pixel contribute more to the classification process, with a larger weight, and vice versa, which supports an improved hyperspectral image classification performance.

The remainder of this paper is organized as follows. Section II introduces the joint sparse representation classification method. Section III proposes the nonlocal weighted joint sparse representation classification algorithm for hyperspectral imagery. The experimental results of the proposed classification algorithm with three hyperspectral images are given in Section IV. Finally, Section V concludes the paper.

II. JOINT SPARSE REPRESENTATION CLASSIFICATION

A. Sparse Representation

In a sparsity model, it is assumed that a signal can be approximated by a sparse linear combination of elements from a basis set. We construct a matrix $\mathbf{A} \in \mathbb{R}^{B \times N}$ with $B \ll N$, and the compact signal $s \in \mathbb{R}^N$ can be approximately represented by multiplying the dictionary \mathbf{A} with a sparse vector α , in which only a few entries colored in grey are non-zero. The sparse vector α can be obtained by solving the following optimization problem:

$$\alpha = \arg \min \|\alpha\|_0 \text{ s.t. } s = \mathbf{A}\alpha + \varepsilon \quad (1)$$

where ε is the model noise vector. It is clear that (1) is a NP-hard combinatorial search problem. Generally, there are two effective ways of solving this problem: the greedy pursuit based algorithms [15], and the ℓ_1 -norm convex relaxation algorithms [16].

B. Sparse Representation Classification (SRC) and Joint Sparse Representation Classification (JSRC)

As an important application of sparse representation, SRC, which was first proposed by Wright *et al.* [10] for face recognition, has been widely used in various pattern recognition applications. For HSI classification, suppose that we have M distinct classes and stack the given N_i ($i = 1, 2, \dots, M$) training pixels from the i th class as columns of a dictionary $\mathbf{A}_i = [\mathbf{a}_{i,1}, \mathbf{a}_{i,2}, \dots, \mathbf{a}_{i,N_i}] \in \mathbb{R}^{B \times N_i}$, where \mathbf{A}_i denotes a linear low-dimensional space, and B refers to the number of bands of the HSI. The hyperspectral pixel \mathbf{s} which belongs to the i th class can be compactly represented as a linear combination of the given training samples, $\mathbf{s} = \mathbf{A}_i\alpha_i + \xi$ and $\alpha_i \in \mathbb{R}^{N_i}$.

As the identity of the signal \mathbf{s} is initially unknown before classification, in order to linearly represent it, we define a new matrix $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_M] \in \mathbb{R}^{B \times N}$, which includes all the training samples, and \mathbf{A}_i is considered as a sub-matrix of \mathbf{A} , where $N = \sum_{i=1}^M N_i$. Signal \mathbf{s} can be described as:

$$\begin{aligned} s &= \mathbf{A}_1\alpha_1 + \dots + \mathbf{A}_i\alpha_i + \dots + \mathbf{A}_M\alpha_M + \xi \\ &= \underbrace{[\mathbf{A}_1 \dots \mathbf{A}_i \dots \mathbf{A}_M]}_{\mathbf{A}} \underbrace{[\alpha_1 \dots \alpha_i \dots \alpha_M]}_{\alpha} + \xi \\ &= \mathbf{A}\alpha + \xi \in \mathbb{R}^B \end{aligned} \quad (2)$$

Ideally, $\alpha \in \mathbb{R}^N$ is a coefficient vector whose entries are all close to zero, except for those associated with the class to which the signal \mathbf{s} belongs. We consider α as an L_0 -sparse vector, where $L_0 = \|\alpha\|_0$.

As the hyperspectral signals from the same class often span the same low-dimensional subspace that is constructed by the corresponding training samples, which involves the non-zero entries of the sparse vector, the class of the hyperspectral signal \mathbf{s} can be directly determined by the characteristics of the recovered sparse vector α . With the sparse approximation result, we classify \mathbf{s} by assigning it to the object class that minimizes the residual r_i :

$$\begin{aligned} \text{class}(s) &= \arg \min_{i=1, \dots, M} r_i(s) \\ &= \arg \min_{i=1, \dots, M} \|\mathbf{s} - \mathbf{A}_i\alpha_i\|_2 \end{aligned} \quad (3)$$

For a hyperspectral image, the neighboring pixels in a small patch often consist of similar materials. All the pixels with similar spectra in the neighborhood can be linearly represented in the same low-dimensional feature subspace \mathbf{A}_i ($i = 1, \dots, M$) with different compact coefficients. With this assumption, Chen *et al.* [11] applied SRC to hyperspectral imagery and proposed a joint sparse representation classification method with the joint sparsity model (JSM), the superiority of which was clearly demonstrated in [17] and [18]. In the JSM model, all the neighboring pixels around the central test pixel are assumed to share a common sparsity pattern with different sets of coefficients. Both the central test pixel and the neighboring pixels are all stacked into the joint signal matrix and sparsely represented by a row-sparse coefficient matrix, which effectively reflects the class property of the central test pixel.

Consider that the single signal \mathbf{s}_p that is located at position p can be sparsely represented by a $B \times N$ structured dictionary \mathbf{A} . Let the neighborhood window size be set as T , and the hyperspectral image patch is stacked to construct the joint signal matrix $\mathbf{S} = [\mathbf{s}_1 \ \mathbf{s}_2 \ \dots \ \mathbf{s}_T]$, with the size of $B \times T$, whose first column is the test pixel located in the center of the hyperspectral image patch, and the rest of the columns are the neighboring pixels around the test pixel. As shown in Fig. 1, it is assumed that all the columns $\{\mathbf{s}_t\}_{t=1, \dots, T}$ share a common sparsity pattern, and then the joint signal matrix can be represented by the JSM:

$$\begin{aligned} \mathbf{S} &= [\mathbf{s}_1 \ \mathbf{s}_2 \ \dots \ \mathbf{s}_T] = [\mathbf{A}\alpha_1 + \xi_1 \ \mathbf{A}\alpha_2 + \xi_2 \ \dots \ \mathbf{A}\alpha_T + \xi_T] \\ &= \mathbf{A} \underbrace{[\alpha_1 \ \alpha_2 \ \dots \ \alpha_T]}_{\Psi} + \Sigma = \mathbf{A}\Psi + \Sigma \end{aligned} \quad (4)$$

where Ψ is the set of all the sparse coefficient vectors $\{\alpha_i\}_{i=1, \dots, T}$, which correspond to all the hyperspectral pixels in the neighborhood window, and are assumed to share the same low-dimensional sub-dictionary with different coefficients. Σ is the model noise matrix corresponding to the joint signal matrix \mathbf{S} . The optimization model corresponding to (4) can be expressed as:

$$\Psi = \arg \min \|\Psi\|_{\text{row},0} \text{ s.t. } \mathbf{S} = \mathbf{A}\Psi + \Sigma \quad (5)$$

where $\|\bullet\|_{\text{row},0}$ denotes the numbers of non-zero rows of Ψ . The optimal solution Ψ is a row-sparse matrix with L_0 non-zero

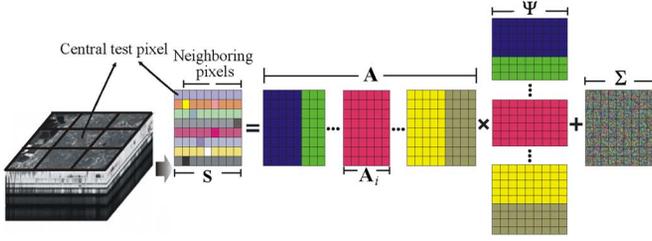


Fig. 1. Joint sparse representation classification (JSRC) of the central test pixel. On the left is a patch of the Moffett Field AVIRIS data from 1997. The top of the cube is a color composite (shown in grayscale) of the three spectral bands. To the right of this is the joint sparse representation flowchart. The input signal has been extended from a 1D vector to a 2D matrix. The first column of the joint signal matrix is the test pixel located in the center of this patch of the image, and the rest of the columns are the neighboring pixels around the test pixel.

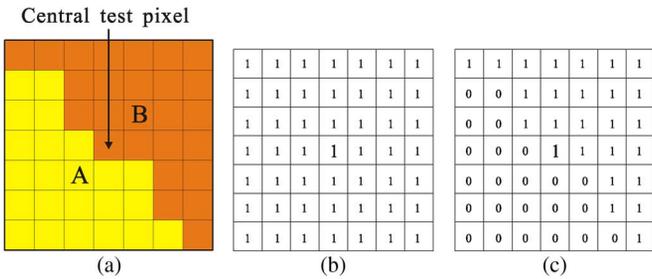


Fig. 2. Weight distributions for classification. (a) Ground truth of a small patch. Pixels in luminous yellow belong to class A, and pixels in ochre yellow belong to class B. (b) The weight distribution for JSM. (c) The ideal distribution for W-JSM.

rows. For the different solutions to the optimization model (5), please refer to [17], [18].

Once the sparse coefficient matrix Ψ is obtained, in a similar way to SRC, we label the central test pixel p of the image patch by the minimal residual:

$$\begin{aligned} class(s_p) &= \arg \min_{i=1, \dots, M} r_i(S) \\ &= \arg \min_{i=1, \dots, M} \|S - A_i \Psi^i\|_F \end{aligned} \quad (6)$$

where Ψ^i denotes the portion of the recovered sparse coefficient matrix Ψ corresponding to the training samples in the i th class.

III. NONLOCAL WEIGHTED JOINT SPARSE REPRESENTATION CLASSIFICATION FOR HYPERSPECTRAL IMAGERY

A. Weighted Joint Sparsity Model

It is noted that in heterogeneous areas, especially around the image edges, neighboring hyperspectral pixels often consist of different materials, which goes against the assumption of the JSM. Therefore, in these cases, it is not fair to take all the neighboring pixels into the joint signal matrix and consider them as equal contributions to the classification process of the central test pixel. We illustrate the limitations of JSRC by taking Fig. 2 as an example.

In order to label the central test pixel, we choose a neighborhood window of size 7×7 for the joint sparse representation classification. Fig. 2(a) denotes the ground truth of the

small patch, where luminous yellow pixels belong to class A, and ochre yellow pixels belong to class B. According to (4), each pixel in the window contributes equally to the classification process of the central pixel, as shown in Fig. 2(b). However, it is obvious that not all the pixels in the neighborhood satisfy the assumption of the joint sparsity model. In fact, the pixels colored in luminous yellow are all dissimilar to the central test pixel and can be regarded as invalid pixels for the classification process. Only the ochre yellow pixels are valid ones. In the ideal case, the contributions of the neighboring pixels should be assigned as shown in Fig. 2(c), to fulfill the assumption of the JSM. That is, the invalid pixels should be discarded in the classification process of the central test pixel.

In order to approach this, we extend the JSM to a weighted joint sparsity model (W-JSM) by weighting all the neighboring pixels of the joint signal matrix. We therefore apply different spatial neighborhood contributions into the joint sparsity model by extending (6) to:

$$\Psi_W = \arg \min \|SW - A\Psi_W\|_F \text{ s.t. } \|\Psi_W\|_{row,0} \leq L_0 \quad (7)$$

where $\mathbf{W} = \text{diag}(w_1, w_2, \dots, w_T)$ is a diagonal matrix, and each entry on its main diagonal denotes the contribution of the corresponding neighboring pixel to the classification process of the central test pixel.

B. Nonlocal Weighting Scheme

Our next task is assigning the appropriate weights to every neighboring pixel of the central test pixel. It is noted that the weight should reflect the correlation between the neighboring pixel and the central test pixel. In this paper, the weights are determined by a nonlocal weighting scheme which aims at explicitly exploiting the self-similarities in images. This approach has been widely used in sparse signal processing [19], [20]. The superiority of the nonlocal weighting scheme over other neighborhood weighting methods is that it not only compares the spectral similarity at the single pixel level, but also the geometrical configuration of the whole neighborhood patch. This weighting scheme addresses the issue of structural preservation to improve the representation performance.

Let the test pixel be denoted as p , and a neighboring pixel is referred to as q . The nonlocal weighting scheme can be mathematically expressed as:

$$w'(p, q) = f(\|J(p) - J(q)\|) \quad (8)$$

where $\|J(p) - J(q)\|$ denotes the similarity measure between the two hyperspectral image patches, which are sized as $so \times so$ and centered at p and q , respectively. $\|J(p) - J(q)\|$ can be expressed as follows:

$$\|J(p) - J(q)\| = \frac{1}{B} \sum_{k=1}^B (I(p)_k - I(q)_k) \otimes \Theta \quad (9)$$

where $I(p)_k$ and $I(q)_k$ represent the digital value vector of the image patches centered at p and q at the k th band, respectively. \otimes is the convolution operator, and Θ is a spatial convolution kernel which measures the weights of the corresponding pixels

in the image patch. The convolution kernel Θ is a Gaussian blur kernel.

$f(\bullet)$ denotes the kernel function, which determines the effectiveness of the nonlocal weighting scheme, one important task of which is to establish an appropriate kernel function for computing the weighting factors, which controls the quality of the image processing result [21]. In this paper, an improved Tukey weight function is employed as the kernel function:

$$\begin{aligned} w'(p, q) &= f(\|J(p) - J(q)\|) \\ &= \left(1 - \left(\frac{\|J(p) - J(q)\|}{\rho}\right)^2\right)^2 \end{aligned} \quad (10)$$

where the parameter ρ is the factor used to adjust the decay of the kernel function. Here, we consider the pixel which satisfies the expression $\max(\|J(p) - J(q)\|)$ as the most dissimilar point to the central test pixel in the neighborhood window, and the $\max(\|J(p) - J(q)\|)$ as the local and adaptive kernel parameter ρ . In this way, we can adaptively calculate the weight of each neighboring pixel in each image patch.

The main aim of our method is to discriminate the valid pixels from the invalid neighboring pixels, as shown in Fig. 2(c). However, the weight we get from (10) is a fractional probability of the contribution that the neighboring pixel offers. Because of the complexity of hyperspectral pixels, it is inappropriate to binarize the weight with a threshold to absolutely discard all the pixels whose weights are smaller than the threshold. We therefore make some modifications to the weighting scheme:

$$w(p, q) = \begin{cases} 0 & 0 < w'(p, q) < w'_1 \\ w'(p, q) & w'_1 < w'(p, q) < w'_2 \\ 1 & w'_2 < w'(p, q) < 1 \end{cases} \quad (11)$$

where w'_1 and w'_2 are the two parameters used to judge valid and invalid neighboring pixels. For the pixels with a weight larger than w'_2 , we regard them as being very similar to the central test pixel, and they are therefore valid pixels. The weights of these pixels are therefore reset to 1. The pixels with weights smaller than w'_1 are considered as highly dissimilar pixels, and are discarded as invalid pixels in the classification process of the central test pixel. The pixels with weights between w'_1 and w'_2 are partially similar to the central test pixel, and their corresponding weights remain unchanged. In this way, when compared with the JSRC method, the impact of the dissimilar pixels is reduced or even eliminated in the classification process of the central test pixel.

Substituting (10) and (11) into (7), we get:

$$\begin{aligned} \Psi_{NLW} &= \arg \min \|SW_{NL} - A\Psi_{NLW}\|_F \\ \text{s.t. } &\|\Psi_{NLW}\|_{row,0} \leq L_0 \end{aligned} \quad (12)$$

where $\mathbf{W}_{NL} = \text{diag}(w(p, q_1), w(p, q_2), \dots, w(p, q_T))$ refers to the nonlocal weighted matrix, and each main diagonal entry of \mathbf{W}_{NL} can be obtained via (10) and (11). The objective function (12) can be solved by a generalized greedy optimization method named simultaneous orthogonal matching

pursuit (SOMP), which is described in detail in [17]. We label the identity of the central pixel \mathbf{s}_p by minimizing the residual:

$$\text{class}(\mathbf{s}_p) = \arg \min_{i=1, \dots, M} \|\mathbf{S}\mathbf{W}_{NL} - A_i(\Psi_{NLW}^i)\|_F \quad (13)$$

where Ψ_{NLW}^i denotes the portion of the recovered nonlocal weighted sparse coefficient matrix Ψ_{NLW} corresponding to the training samples in the i th class.

C. Procedure of NLW-JSRC

By incorporating the nonlocal spatial structure information, the implementation details of the proposed NLW-JSRC algorithm are summarized in **Algorithm 1**.

IV. EXPERIMENTS

In this section, we investigate the effectiveness of the proposed NLW-JSRC algorithm with three hyperspectral images. The classical classifiers of SVM with radial basis function (RBF) kernel [8], OMP-S (OMP with Laplacian smoothing) [11], and JSRC (referred to as SOMP in [11]) are used as benchmarks in this paper. In addition, the OMP method and the homotopy algorithm [22] for ℓ_1 -norm convex relaxation, which are used to solve the SRC problems, are also included in the comparisons. We set the size of the nonlocal weighting patch as $so \times so = 7 \times 7$, as recommended in the work of [13]. The weighting factor for OMP-S is fixed to 1, as recommended in [11]. We set $w'_1 = 0.14$ and $w'_2 = 0.88$ for the final proposed NLW-JSRC algorithm. For the JSRC and NLW-JSRC algorithms, the sparsity level L_0 is chosen between $L_0 = 5$ and $L_0 = 80$, and the neighborhood window size is chosen between 3×3 ($T = 9$) and 13×13 ($T = 169$). All the experiments were conducted using MATLAB R2011b on a 3.50 GHz machine with 8.0 Gb RAM.

Algorithm 1. The NLW-JSRC procedure for HSI.

1. **Input:** 1) A dictionary of training samples $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_M] \in \mathbb{R}^{B \times N}$ for M classes; 2) A data set, in which each pixel located at p is represented as $\mathbf{s}_p \in \mathbb{R}^B$; 3) The neighborhood window size T and the nonlocal image patch size $so \times so$
 2. **Initialization:** Construct the dictionary \mathbf{A} with the training samples in the hyperspectral image, and normalize the columns of \mathbf{A} to have unit ℓ^2 -norm
 3. **For** each pixel p in the hyperspectral image:
 - 1) Construct the initial joint signal matrix $\mathbf{S} = [\mathbf{s}_1 \ \mathbf{s}_2 \ \dots \ \mathbf{s}_T] \in \mathbb{R}^{B \times T}$, where p locates at the center of the neighborhood window, and normalize the columns of \mathbf{S} to have unit ℓ^2 -norm
 - 2) Calculate the nonlocal weight matrix \mathbf{W}_{NL} , and set $\mathbf{S}_{NLW} = \mathbf{S}\mathbf{W}_{NL}$
 - 3) Calculate the sparse coefficient matrix Ψ_{NLW} from (12)
 - 4) Compute the residuals and label the test pixel p by (15)
 - 5) Turn to the next test pixel
 - End For
 4. **Output:** A 2-D matrix which records the labels of the all pixels
-

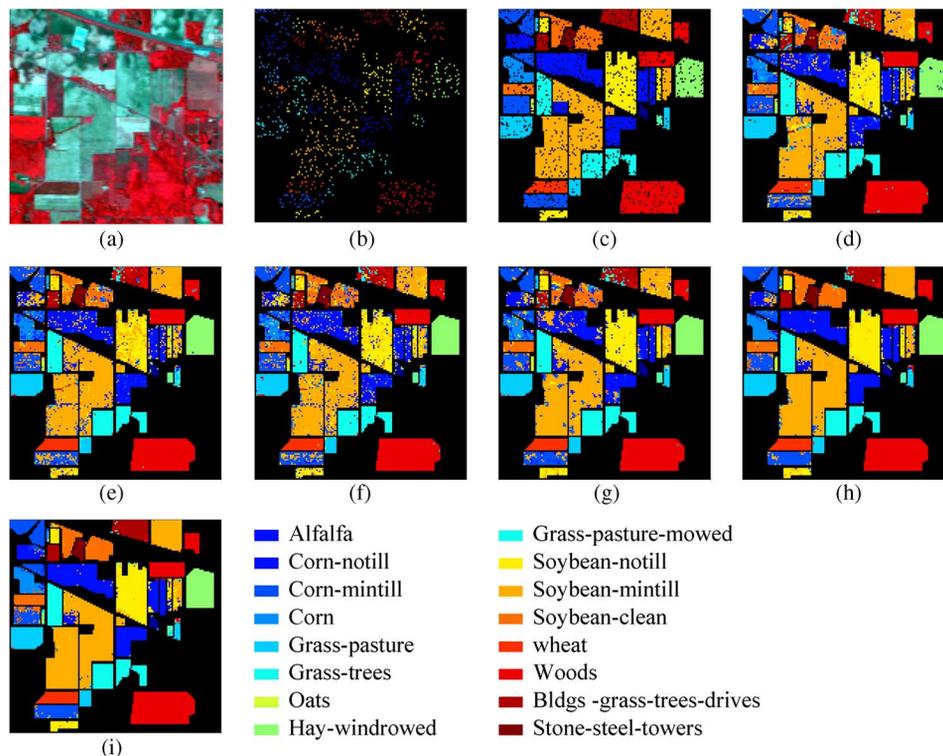


Fig. 3. Classification results with the Indian Pines image: (a) false color image (R:57, G:27, B:17), (b) training set, (c) test set, (d) SVM, (e) ℓ_1 , (f) OMP, (g) OMP-S, (h) JSRC, (i) NLW-JSRC.

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 northwestern Indiana, and consists of 145×145 pixels and 224 spectral reflectance bands in the wavelength range $0.4\text{--}2.5 \mu\text{m}$. The false color composite of the Indian Pines image is shown in Fig. 3(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 [23]. In the reference data with 16 classes, we randomly sampled 9% of the labeled data as the training samples, and the remainder as the test samples. The numbers of the training and test sets are shown in Table I, and the visual maps are shown in Fig. 3(b) and (c), respectively. We stack all the training samples as columns of the dictionary $\mathbf{A} \in \mathbb{R}^{200 \times 958}$ in every sparsity-based algorithm for this data set.

The optimal parameter settings for the proposed NLW-JSRC were $L_0 = 30$ and $T = 81$, while the corresponding optimal parameters for JSRC were $L_0 = 20$ and $T = 25$. The classification maps of the various classification methods are shown in Fig. 3(d)–(i), respectively. The quantitative evaluation results, which include the classification accuracies for every class, the overall accuracy, and the kappa coefficient, are shown in Table II. It can be seen that, for most classes, the three sparsity-based algorithms with spatial information (OMP-S, JSRC, and NLW-JSRC) outperform the others, and the NLW-JSRC algorithm shows the best performance. It can be seen that by incorporating the nonlocal spatial information, the proposed NLW-JSRC algorithm leads to a better classification map than JSRC.

TABLE I
THE 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

We next demonstrate the impact of the two parameters w'_2 and w'_1 on the classification result of the Indian Pines image. We conducted 64 experiments by ranging w'_2 from 0.76 to 0.9, and w'_1 from 0.1 to 0.24. The curved surface in yellow shows the result of the proposed NLW-JSRC method, and the blue surface which is parallel to the horizontal plane refers to that of JSRC, as shown in Fig. 4. It can be clearly seen that the setting of w'_2 and w'_1 has a certain positive effect on the classification result of NLW-JSRC, which outperforms the JSRC method for any parameter setting within an appropriate range.

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

Class	SVM	ℓ_1	OMP	OMP-S	JSRC	NLW-JSRC
1	0.7250	0.8250	0.8000	0.9750	0.9250	0.9500
2	0.7698	0.7706	0.7167	0.7721	0.9145	0.9299
3	0.7577	0.6760	0.7149	0.7807	0.8527	0.8782
4	0.6056	0.6291	0.5775	0.6385	0.8122	0.8545
5	0.9080	0.8552	0.8966	0.9609	0.9172	0.9333
6	0.9696	0.9848	0.9680	0.9939	1	1
7	0.9130	0.5217	0.7826	1	0.7826	0.7391
8	0.9605	0.9977	0.9977	1	1	1
9	0.6875	0.6875	0.7500	1	0.3125	0.3125
10	0.7691	0.6274	0.7166	0.8971	0.8766	0.9051
11	0.8145	0.8641	0.7928	0.8709	0.9464	0.9690
12	0.6423	0.6479	0.6423	0.7266	0.9195	0.9682
13	0.9130	1	0.9891	0.9946	1	1
14	0.9548	0.9861	0.9513	0.9809	0.9878	0.9991
15	0.5331	0.5360	0.5994	0.6568	0.8963	0.9625
16	0.9259	0.9877	0.9506	0.9259	1	0.9753
OA	0.8163	0.8154	0.7995	0.8678	0.9313	0.9519
κ	0.7901	0.7873	0.7708	0.8488	0.9215	0.9450

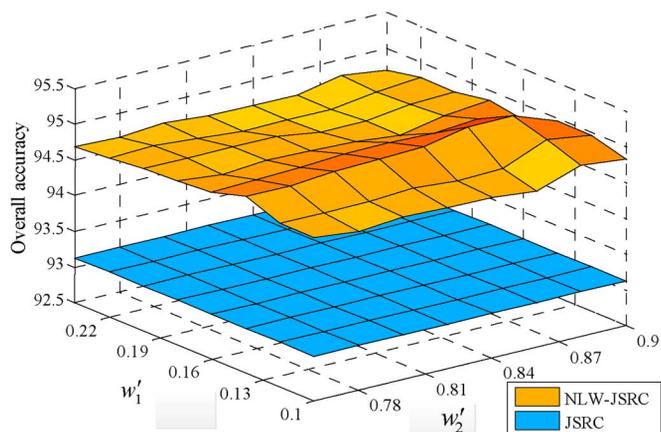


Fig. 4. The impact of the parameters w'_1 and w'_2 on the Indian Pines image.

We then demonstrate the effects of the neighborhood size T and sparsity level L_0 on the performances of JSRC and NLW-JSRC. In Fig. 5, the size of the neighborhood ranges from 3×3 ($T = 9$) to 13×13 ($T = 169$). The horizontal axis indicates the neighborhood size T , and the vertical axis shows the corresponding optimal overall accuracy (%) of the different neighborhood sizes. It can be seen that the classification result of the proposed method is slightly better than that of JSRC for most neighborhood sizes. After reaching the extreme point, the accuracy curve of the proposed method is more stable than that of the JSRC method. It can also be concluded that NLW-JSRC is more robust than JSRC with regard to different neighborhood sizes. Fig. 6 shows the optimal results under different L_0 sparsity levels, which range from 5 to 80. Both the plots rise quickly and reach a maximum point, then remain relatively stable with only a tiny decline. It can be seen that the proposed NLW-JSRC algorithm outperforms JSRC at most sparsity levels.

Table III shows the running times for the sparsity-based algorithms when reaching their optimal classification performances. For the single signal oriented algorithms, the running time for the greedy pursuit based algorithm (i.e., OMP) is faster

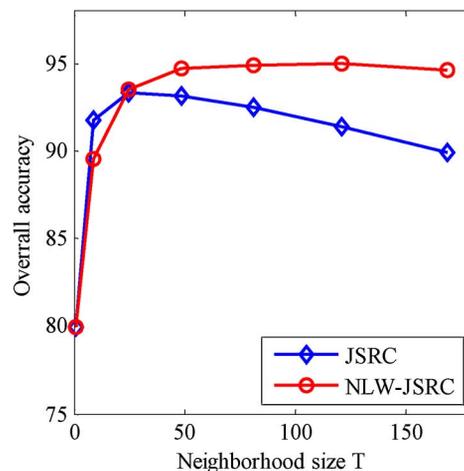


Fig. 5. The classification results versus different neighborhood sizes T .

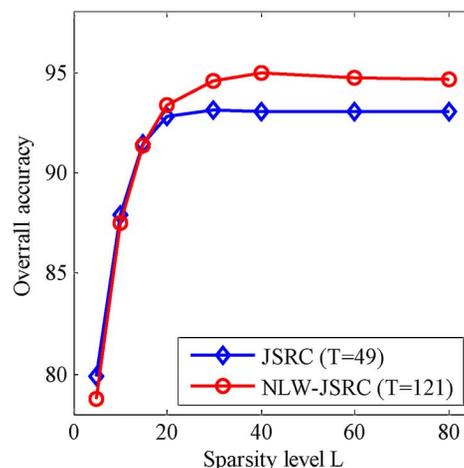


Fig. 6. The classification results versus different sparsity levels L_0 .

TABLE III
SPEED WITH THE INDIAN PINES IMAGE FOR THE SPARSITY-BASED ALGORITHMS

	ℓ_1	OMP	OMP-S	JSRC	NLW-JSRC
Times(s)	134.0535	27.3001	214.2989	61.3195	410.3501

than the ℓ_1 -norm convex relaxation algorithm, as the latter is far more sophisticated than the greedy algorithms [24]. For the three sparsity-based algorithms with spatial information, it can be seen that JSRC is the fastest, and the proposed algorithm requires more computing time, due to the nonlocal weighting scheme increasing the computational load. It is, however, reasonable to believe that with the rapid development in computer hardware, the time cost of the proposed method will no longer be an issue.

B. ROSIS Urban Data: Pavia University, Italy

This scene was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor during a flight campaign over Pavia University, northern Italy. The number of spectral bands is 103 and the image size is 610×610 pixels. We cut a patch sized 610×340 from the original image. The geometric

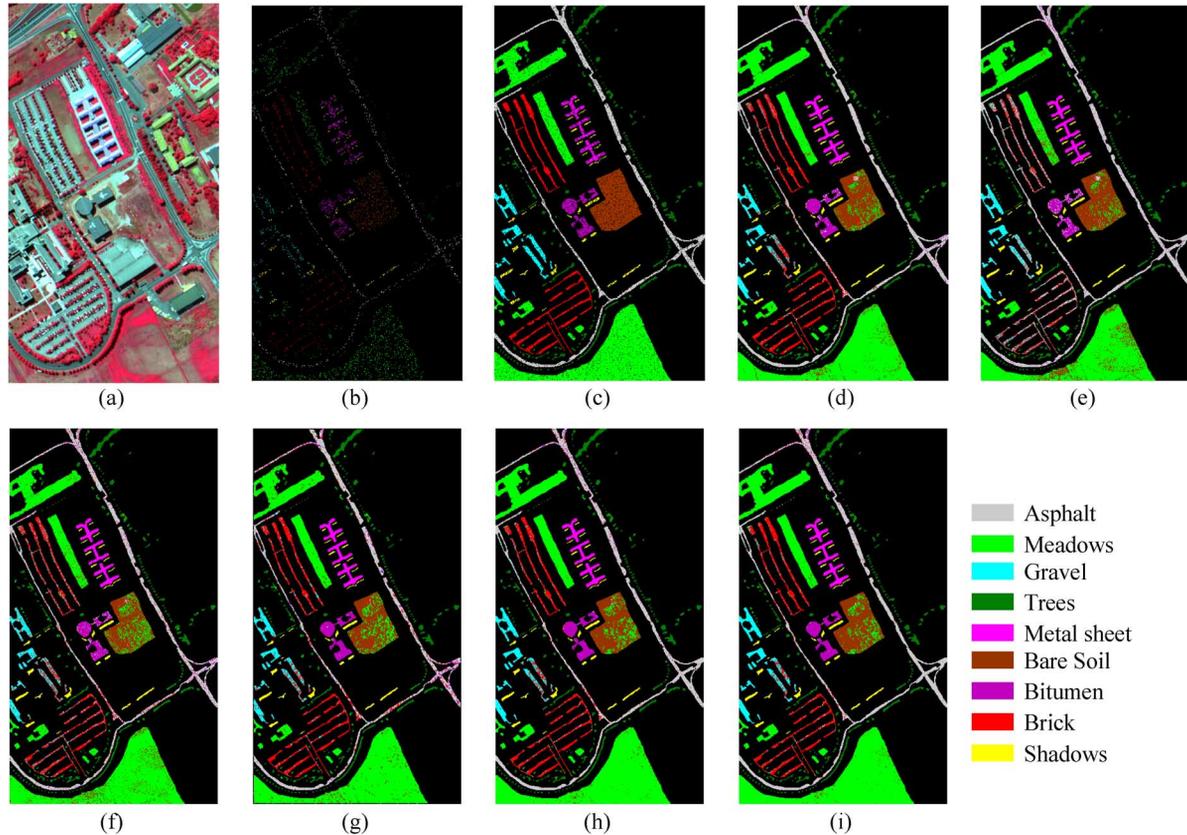


Fig. 7. Classification results of the Pavia University image: (a) false color image (R:102, G:56, B:31), (b) training set, (c) test set, (d) SVM, (e) ℓ_1 , (f) OMP, (g) OMP-S, (h) JSRC, (i) NLW-JSRC.

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

No.	Class name	Samples	
		Train	Test
1	Asphalt	597	6034
2	Meadows	932	17717
3	Gravel	189	1910
4	Trees	276	2788
5	Metal sheet	269	1076
6	Bare Soil	453	4576
7	Bitumen	266	1064
8	Brick	331	3351
9	Shadows	189	758
Total		3502	35204

TABLE V
CLASSIFICATION ACCURACY (%) FOR THE PAVIA UNIVERSITY IMAGE WITH THE TEST SET

Class	SVM	ℓ_1	OMP	OMP-S	JSRC	NLW-JSRC
1	0.9069	0.9165	0.8016	0.7539	0.8642	0.8767
2	0.9585	0.8970	0.9189	0.9609	0.9867	0.9891
3	0.7063	0.7429	0.6513	0.7521	0.8063	0.7942
4	0.9358	0.9494	0.9085	0.9147	0.9412	0.9290
5	0.9963	0.9991	0.9972	0.9991	1	1
6	0.8059	0.8453	0.7193	0.7354	0.7867	0.7769
7	0.8703	0.7566	0.8158	0.9530	0.9624	0.9643
8	0.8729	0.3772	1	0.8821	0.8526	0.8568
9	0.9987	0.9855	0.7129	0.9565	0.9842	0.9881
OA	0.9110	0.8466	0.8463	0.8829	0.9208	0.9298
κ	0.8799	0.7954	0.7933	0.8423	0.8924	0.9046

resolution is 1.3 m. The false color composite of the Pavia University image is shown in Fig. 7(a). As Table IV shows, this image contains nine ground-truth classes. We randomly sampled 9% of the data in each class as the training samples and the remainder as the test samples. The training and test sets are visually shown in Fig. 7(b) and (c). All the training samples are stacked as the columns of the dictionary $\mathbf{A} \in \mathbb{R}^{103 \times 3502}$ in every sparsity-based algorithm for this data set.

The optimal parameter settings for the proposed NLW-JSRC method were $L_0 = 15$ and $T = 25$, and for JSRC they were $L_0 = 10$ and $T = 9$. For the high spatial resolution of this image, even pixels belonging to the same class in a spatial window may show some variance, which results in the optimal spatial window size being smaller than in the experiments in

Section IV-A. The classification results for the various different classifiers are visually displayed in Fig. 7(d)–(i), respectively. The quantitative evaluation results, which include the classification accuracies for each class, the overall accuracy, and the kappa coefficient, are shown in Table V. It can be seen that the proposed NLW-JSRC method yields the best overall accuracy, kappa coefficient, and classification accuracy for most classes.

For the Pavia University data set with high spatial resolution, the optimal neighborhood size tends to be smaller than with the last data set, and the running time of the JSRC method is close to that of the OMP method. It is, however, still observed that the proposed NLW-JSRC method spends the most time, with more computational burden.

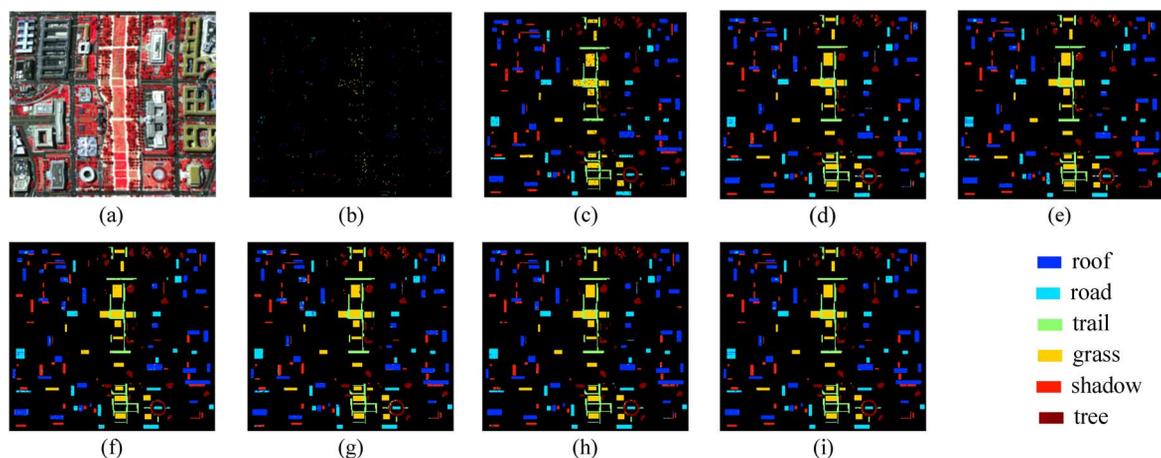


Fig. 8. Classification results of the Washington DC image: (a) false color image (R:63, G:52, B:36), (b) training set, (c) test set, (d) SVM, (e) ℓ_1 , (f) OMP, (g) OMP-S, (h) JSRC, (i) NLW-JSRC.

TABLE VI
SPEED WITH THE PAVIA UNIVERSITY IMAGE FOR
THE SPARSITY-BASED ALGORITHMS

	ℓ_1	OMP	OMP-S	JSRC	NLW-JSRC
Times(s)	1391.6111	220.5124	1731.8445	220.9821	1902.9045

TABLE VII
THE SEVEN 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

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, who 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 $\times 10 \mu\text{m}$ region, 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. 8(a). This image contains six ground-truth classes. We randomly sampled around 5% of the labeled pixels in each class for training, and the rest were used for testing. The training and test sets are visually shown in Fig. 8(b) and (c), respectively. In addition, the numbers of training and test samples are also shown in Table VII. The size of the dictionary \mathbf{A} for this experiment was 191×495 .

TABLE VIII
CLASSIFICATION ACCURACY (%) FOR THE WASHINGTON DC MALL IMAGE
WITH THE TEST SET

Class	SVM	ℓ_1	OMP	OMP-S	JSRC	NLW-JSRC
1	0.9273	0.9802	0.9643	0.9273	0.9839	0.9805
2	0.9512	0.9272	0.9580	0.9512	0.9962	0.9797
3	0.9367	0.9883	0.9883	0.9367	0.9842	0.9917
4	0.9641	0.9900	0.9924	0.9641	0.9941	0.9953
5	0.9173	0.9981	0.9709	0.9173	0.8741	0.9671
6	0.9585	0.9550	0.9797	0.9586	0.9841	0.9815
OA	0.9585	0.9745	0.9742	0.9412	0.9751	0.9831
κ	0.9483	0.9682	0.9679	0.9270	0.9690	0.9790

TABLE IX
SPEED WITH THE WASHINGTON DC MALL IMAGE FOR
THE SPARSITY-BASED ALGORITHMS

	ℓ_1	OMP	OMP-S	JSRC	NLW-JSRC
Times(s)	72.9894	22.7561	98.6113	10.4438	124.9296

The optimal parameter settings for the NLW-JSRC method were $L_0 = 15$ and $T = 9$. The corresponding optimal parameters for JSRC were $L_0 = 20$ and $T = 9$. The classification results are shown in Fig. 8(d)–(i), respectively. The quantitative evaluation results, which include the classification accuracies for every class, the overall accuracy, and the kappa coefficient, are shown in Table VIII. The running times of every sparsity-based algorithm are given in Table IX. From Tables VIII and IX, it can be seen that the proposed NLW-JSRC method yields the best overall performance, but is slower than the others.

V. CONCLUSION

In this paper, we propose a new nonlocal weighted joint sparse representation classification (NLW-JSRC) method for hyperspectral imagery, to support improved classification capabilities. In the classification process of the central test pixel, different neighboring pixels are assigned different weights, which are determined by the structural similarity between the neighboring pixels and the central test pixel. The nonlocal weighted joint sparsity model is solved by the simultaneous

orthogonal matching pursuit (SOMP) technique. The main advantage of the proposed NLW-JSRC method is that the nonlocal weighted joint sparsity model exploits the different contributions of the neighboring pixels to the classification process of the central test pixel by incorporating nonlocal spatial structure information. The proposed nonlocal weighted joint sparse representation classification method was tested on three hyperspectral images. The extensive experimental results clearly indicate that the proposed NLW-JSRC method can achieve competitive classification results.

However, the proposed algorithm still has room for improvement. For example, other improved weighting methods with a reduced computational burden could be introduced to improve the computational efficiency. Furthermore, another task on our agenda is to seek the optimal neighborhood window size for each pixel adaptively in the whole image. These issues will be the main focus of our future work.

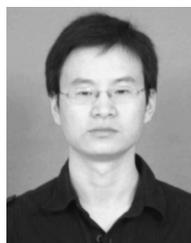
ACKNOWLEDGMENT

The authors would like to thank Prof. D. Landgrebe from Purdue University for providing the AVIRIS image of Indian Pines and the HYDICE image of Washington DC Mall, Prof. Gamba from University of Pavia for providing the ROSIS dataset and Dr. J. Mairal from INRIA for sharing the SPAMS software package. Thanks also to the associate editor and the anonymous reviewers for their careful reading and helpful remarks.

REFERENCES

- [1] E. Christophe, D. Leger, and C. Mailhes, "Quality criteria benchmark for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 9, pp. 2103–2114, Sep. 2005.
- [2] P. K. Goel, S. O. Prasher, R. M. Patel, J. A. Landry, R. B. Bonnell, and A. A. Viau, "Classification of hyperspectral data by decision trees and artificial neural networks to identify weed stress and nitrogen status of corn," *Comput. Electron. Agr.*, vol. 39, no. 2, pp. 67–93, May 2003.
- [3] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.
- [4] J. A. Palmason, J. A. Benediktsson, J. R. Sveinsson, and J. Chanussot, "Classification of hyperspectral data from urban areas using morphological preprocessing and independent component analysis," in *Proc. 2005 Int. Conf. Geoscience and Remote Sensing Symp. (IGARSS)*, pp. 25–29.
- [5] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. 5th Annual Workshop on Computational Learning Theory*, 1992, pp. 144–152.
- [6] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer Verlag, 1999.
- [7] I. W. Tsang, J. T. Kwok, and P. M. Cheung, "Core vector machines: Fast SVM training on very large data sets," *J. Mach. Learn. Res.*, vol. 6, p. 363, 2006.
- [8] L. Bruzzone, C. Mingmin, and M. Marconcini, "A novel transductive SVM for semisupervised classification of remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3363–3373, Nov. 2006.
- [9] H. Zhang, A. C. Berg, M. Maire, and J. Malik, "SVM-KNN: Discriminative nearest neighbor classification for visual category recognition," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVPR)*, 2006, pp. 2126–2136.
- [10] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.
- [11] 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.
- [12] D. Baron, M. F. Duarte, M. B. Wakin, S. Sarvotham, and R. G. Baraniuk, "Distributed Compressive Sensing," Arxiv preprint arXiv:0901.3403, 2009.

- [13] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVPR)*, 2005, pp. 60–65.
- [14] A. A. Efros and T. K. Leung, "Texture synthesis by non-parametric sampling," in *Proc. 1999 IEEE Int. Conf. Computer Vision (ICCV)*, vol. 2, pp. 1033–1038.
- [15] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
- [16] D. L. Donoho and Y. Tsaig, "Fast Solution of l_1 -Norm Minimization Problems When the Solution May be Sparse," Dept. Statistics, Stanford Univ., Stanford, CA, USA, 2006.
- [17] 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, May 2006.
- [18] J. A. Tropp, "Algorithms for simultaneous sparse approximation. Part II: Convex relaxation," *Signal Process.*, vol. 86, pp. 589–602, 2006.
- [19] 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.
- [20] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proc. 2009 IEEE Int. Conf. Computer Vision (ICCV)*, vol. 2, pp. 2272–2279.
- [21] J. Tian, W. Yun, and S. Xie, "On the kernel function selection of non-local filtering for image denoising," in *Proc. 2008 IEEE Int. Conf. Machine Learning and Cybernetics*, pp. 2964–2969.
- [22] A. Y. Yang, S. S. Sastry, A. Ganesh, and Y. Ma, "Fast l_1 -minimization algorithms and an application in robust face recognition: A review," in *Proc. 17th IEEE Int. Conf. Image Processing (ICIP)*, 2010, pp. 1849–1852.
- [23] J. A. Gualtieri and R. F. Crompton, "Support vector machines for hyperspectral remote sensing classification," in *27th AIPR Workshop: Advances in Computer-Assisted Recognition*, 1999, pp. 221–232.
- [24] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. New York, NY, USA: Springer Verlag, 2010.



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.



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, China.

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



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 Lab of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China, in 2010.

He is currently a Lecturer with the Department of Survey Engineering, Xi'an University of Science and Technology, Xi'an, China. 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 of 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 "Chang-Jiang Scholar" Chair Professor appointed by the Ministry of Education, China. He has more than 260 research papers and 5 patents. He is now Principal Scientist for the China State Key Basic Research Project (2011–2016) appointed by the Ministry of National

Science and Technology of China to lead the remote sensing program in China. His research interests include hyperspectral remote sensing, high resolution remote sensing, image processing and artificial intelligence.

Prof. Zhang regularly serves as a Co-Chair of the series SPIE Conferences on Multispectral Image Processing and Pattern Recognition (MIPPR), Conference on Asia Remote Sensing, and many other conferences. He edits several conference proceedings, issues, and the Geoinformatics Symposiums. He also serves as an Associate Editor of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, *International Journal of Ambient Computing and Intelligence* (IJACI), *International Journal of Image and Graphics*, *International Journal of Digital Multimedia Broadcasting*, *Journal of Geo-spatial Information Science*, and *Journal of Remote Sensing*. He is a Fellow of IEE, Executive Member (Board of Governors) of the China National Committee of International Geosphere-Biosphere Programme, and Executive Member for the China Society of Image and Graphics.