

# Joint Collaborative Representation With Multitask Learning for Hyperspectral Image Classification

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

**Abstract**—In this paper, we propose a joint collaborative representation (CR) classification method with multitask learning for hyperspectral imagery. The proposed approach consists of the following aspects. First, several complementary features are extracted from the hyperspectral image. We next apply these features into the unified multitask-learning-based CR framework to acquire a representation vector and adaptive weight for each feature. Finally, the contextual neighborhood information of the image is incorporated into each feature to further improve the classification performance. The experimental results suggest that the proposed algorithm obtains a competitive performance and outperforms other state-of-the-art regression-based classifiers and the classical support vector machine classifier.

**Index Terms**—Classification, hyperspectral imagery (HSI), joint collaborative representation (CR) model, multitask learning, sparse representation.

## I. INTRODUCTION

**H**YPERSPECTRAL imagery (HSI), spanning the visible to the infrared spectrum with a high spectral resolution, provides fine spectral differences between various materials of interest, thus supporting improved target detection and classification capabilities, relative to multispectral images [1], [2]. In the hyperspectral supervised classification case, the class label of each pixel, denoted by a vector whose entries correspond to the narrow spectral band responses, is determined by a given training set from each class [3]–[6]. The support vector machine (SVM) [7]–[11], a supervised learning method that generates input–output mapping functions from a set of labeled training data, is a powerful HSI classifier [12]–[15]. Melgani and Bruzzone [12] assessed the potential of the SVM classifier in the hyperdimensional feature spaces, with a pleasing classification performance. Fauvel *et al.* [13] incorporated spatial information into the spectral-only SVM classifier by the fusion of the morphological information and the original hyperspectral data.

Manuscript received February 6, 2013; revised June 6, 2013 and October 15, 2013; accepted November 30, 2013. 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 41101336, and by the Program for Changjiang Scholars and Innovative Research Team in University IRT1278.

J. Li, H. Zhang, L. Zhang, and X. Huang are with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: zjjerica@163.com; zhanghongyan@whu.edu.cn; zlp62@whu.edu.cn; huang\_wlu@163.com).

L. Zhang is with the School of Computer, Wuhan University, Wuhan 430072, China (e-mail: zhanglefei@whu.edu.cn).

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

Digital Object Identifier 10.1109/TGRS.2013.2293732

Pal and Foody [14] noted that feature selection could be a valuable preprocessing operation for the SVM classification method. Tarabalka *et al.* [15] presented an accurate spectral–spatial-based probabilistic SVM pixelwise method by adding the use of a Markov random field regularization. Recently, the collaborative representation (CR) mechanism [16], which approximates the unlabeled pixel as a linear combination of all of the training samples, has also played an important role in HSI classification [17]–[19].

In recent years, it has become common to design a classifier which combines multiple features (e.g., spectral, texture, and shape) to improve the classification accuracy for HSI [20], [21]. The idea behind this is that one single feature can only depict the hyperspectral image from one perspective, and it is obvious that none of the common feature descriptors will have the same discriminative power for all classes [22]. To effectively characterize hyperspectral data, it is natural to combine several complementary feature descriptions to represent a pixel's information more fully. Zhang *et al.* [23] introduced a patch alignment framework to linearly combine multiple features and obtain a unified low-dimensional representation of these multiple features for the subsequent SVM classification. Jimenez *et al.* [24] focused on the different influences of various linear unsupervised feature extraction methods for high-dimensional data classification. Zheng *et al.* [25] applied the multitask joint sparse representation classification (MTJSRC) proposed in [26] with spatial filtering postprocessing into large-scale satellite image annotation and obtained excellent results.

In HSI, neighboring pixels usually consist of similar materials (i.e., highly spectrally correlated pixels), which can be used in the classification procedure to support the analysis performance [13], [15]. Waqas *et al.* [27] pruned the over-complete dictionary to an automatically chosen optimal nearest basis with the assumption of locally linear embedding to steady the performance of the CR-based classifier. Chen *et al.* [17] proposed a joint sparse representation classification method (JSRC) based on the joint sparsity model (JSM) [28] with neighborhood pixel information to support HSI classification.

In view of this, we combine multiple features and the spatial neighborhood information into a unified framework and propose a joint CR classification method with multitask learning (JCRC-MTL) for HSI. The contributions of this paper are twofold. First, multiple features are integrated via a unified multitask learning approach to complement each other for pixel recognition. Second, the neighborhood pixels around the unlabeled pixel are constructed as a joint signal set and are integrated in the multitask joint CR framework to utilize

the neighborhood information. The proposed method aims at enhancing the discrimination of hyperspectral pixels by combining the complementary information of different features in a multitask learning fusion to achieve an improved classification result. Experiments on several hyperspectral images validated the effectiveness of the proposed JCRC-MTL algorithm.

We next discuss the relationship between the proposed JCRC-MTL classification method and the other CR-related methods. By stacking the training samples from all of the classes simultaneously to construct the overcomplete dictionary, it is assumed that every pixel belongs to one of these classes and it will be located in some low-dimensional subspace of the feature spaces spanned by the dictionary. In view of this, Wright *et al.* [29] first applied sparse representation with an  $\ell_1$ -norm regularization to the pattern recognition task and successfully explored the discriminative ability of the sparse coefficient vector. Some further studies based on this work have also been developed. Under the assumption of the JSM, Chen *et al.* [17] utilized the contextual information and constrained the representation coefficient of the joint signal matrix constructed by stacking the neighborhood pixels around the central unlabeled one and itself with the row-sparse  $\ell_{\text{row},0}$ -norm, which denotes the number of nonzero rows of the variable, to achieve a better classification result. Yuan *et al.* [26] and Zheng *et al.* [25] both realized the complementary nature of multiple features and built multitask joint sparse representation models which exploit the invariant similarities between various features. It has been observed that the algorithms mentioned previously can achieve a better classification result than the classical SRC [29]. Although reaching an improved recognition rate, it is still a factor that the sparsity constraint with  $\ell_1$ -norm is time-consuming. In recent years, Zhang *et al.* [16] further revealed that it is not necessary to regularize the sparse coefficient  $\alpha$  by the computationally expensive  $\ell_1$ -norm regularization when the feature dimension is high enough. They pointed out that it is the CR, which represents the test pixel collaboratively with training samples from all of the classes, rather than the sparsity constraint, that determines the performance of the algorithm. In the representation stage, the CR classification via regularized least squares (referred to as CRC in this paper) used in [16] just takes the  $\ell_2$ -norm instead of the  $\ell_1$ -norm.

With the contextual prior, we propose a joint CR model for HSI classification, which is similar to the JSM but with much less computational complexity [30]. In this way, the CRC can be considered as a special case of JCRC, with only one pixel in the joint signal input. Moreover, Yang *et al.* [31] investigated the distinctiveness and similarity of the different features in the multiple feature model in the CR-based framework and achieved a state-of-the-art performance. Since these priors, as well as the sparsity constraint, have been shown to perform well, we integrate the aforementioned techniques and use the multiple features and contextual neighborhood information to further improve the classification performance.

The rest of this paper is organized as follows. Section II briefly reviews the related works. Section III proposes the joint CR classification method with multitask learning (JCRC-MTL) for HSI. The experimental results of the proposed algorithm are given in Section IV. Finally, Section V concludes this paper.

## II. CR WITH MULTIPLE FEATURES FOR HSI

In this section, we first introduce four typical image features for HSI and then introduce a CR classification method that integrates the four features with a weighted multitask learning method.

### A. Multiple Feature Extraction

In the proposed algorithm, four image features are utilized. Each feature of a pixel is represented as a single vector. The four features comprise the spectral value feature, the spectral gradient feature, the Gabor texture feature, and the differential morphological profile (DMP). Examples of these four features extracted from a Reflective Optics System Imaging Spectrometer (ROSIS) image over Pavia city, Italy, are illustrated in Fig. 1.

1) *Spectral Value Feature*: The spectral feature of a hyperspectral pixel is acquired by arranging the digital number (DN) of all of the  $l$  bands

$$\mathbf{s}_{\text{Spectral}} = [s_1, s_2, \dots, s_i, \dots, s_l]' \in \mathbb{R}^l \quad (1)$$

where  $s_i$  denotes the DN in band  $i$ .

2) *Spectral Gradient Feature*: The spectral gradient is a surface reflectance descriptor which is invariant to scene geometry and incident illumination [33]. This feature encodes the information at discrete spectral locations about how fast the surface albedo changes as the spectrum changes. The spectral gradient feature of a hyperspectral pixel is calculated as follows:

$$\begin{aligned} \mathbf{s}_{\text{Gradient}} &= [g_1, g_2, \dots, g_{i-1}, \dots, g_{l-1}]' \\ &= [s_2 - s_1, s_3 - s_2, \dots, s_i \\ &\quad - s_{i-1}, \dots, s_l - s_{l-1}]' \in \mathbb{R}^{l-1} \end{aligned} \quad (2)$$

where  $s_i$  denotes the spectral value of band  $i$  and  $g_{i-1}$  is the  $i-1$ th entry of the gradient vector.

3) *Gabor Texture Feature*: The Gabor wavelet filter [34], whose impulse response is defined by a Gaussian envelope and a complex plane wave, has been widely used in HSI analysis [35], [36]. In this paper, we perform a 2-D Gabor wavelet transform corresponding to the orientation and scale of the physical structures on the first principal component (PC), denoted as  $I$  of the hyperspectral image [32], to extract the Gabor texture feature.

The generalized 2-D Gabor function can be defined as [23]

$$G_{s,d}(x, y) = G_{\bar{\mathbf{v}}}(\bar{\mathbf{x}}) = \frac{\|\bar{\mathbf{v}}\|}{\delta^2} \cdot e^{-\frac{\|\bar{\mathbf{v}}\|^2 \cdot \|\bar{\mathbf{x}}\|^2}{2\delta^2}} \cdot \left[ e^{i \cdot \bar{\mathbf{v}} \cdot \bar{\mathbf{x}}} - e^{-\frac{\delta^2}{2}} \right] \quad (3)$$

where  $\bar{\mathbf{x}} = (x, y)$  is the spatial domain variable and the frequency vector  $\bar{\mathbf{v}}_{(s,d)} = (\pi/2f^s) \cdot e^{i(\pi d/8)}$ , in which  $f = 2$ ,  $s = 0, 1, \dots, 4$  and  $d = 0, 1, \dots, 11$ , which determines the 5 scales and the 12 directions of the Gabor function. The number of oscillations under the Gaussian envelope is determined by  $\delta = 2\pi$ . The Gabor texture feature contains the magnitude information in the first PC  $I$  with the Gabor function of the specific scale parameter  $s$  and direction parameter  $d$

$$F_{s,d}(x, y) = G_{s,d}(x, y) * I(x, y). \quad (4)$$

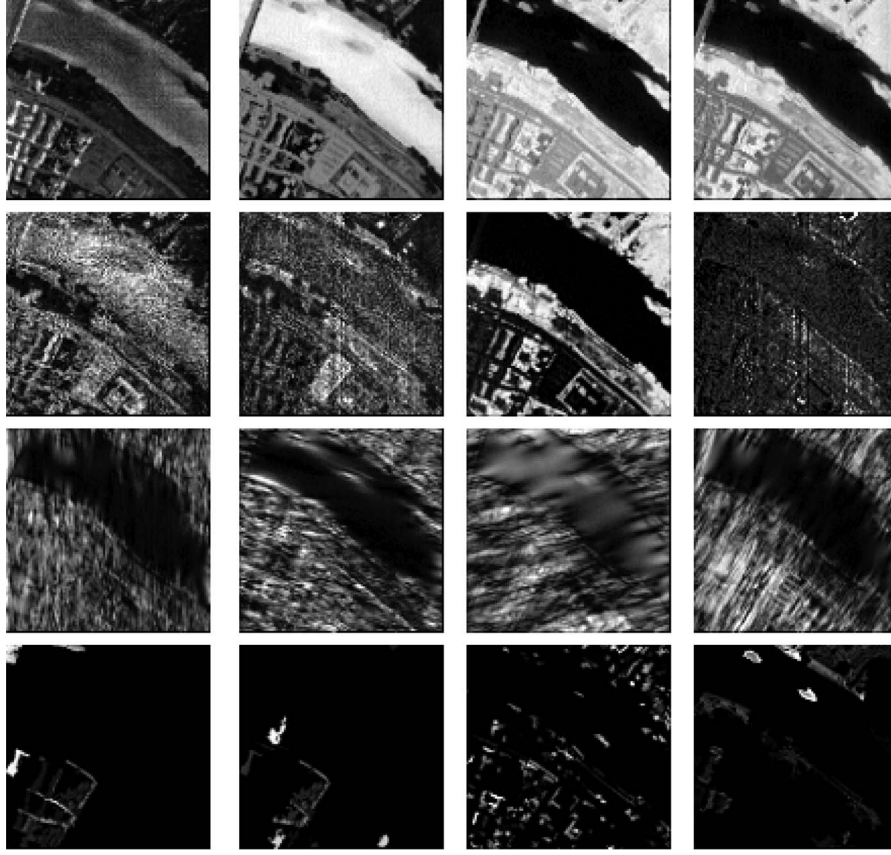


Fig. 1. Some bands of the four features of a ROSIS image over Pavia city, Italy. First row: the spectral value feature images in spectral bands 3, 36, 78, and 102. Second row: the spectral gradient feature images in dimensions 20, 45, 74, and 88. Third row: the Gabor texture feature images in dimensions 2, 26, 40, and 52 extracted from the first PC of the hyperspectral image [32]. Bottom row: the DMP shape feature in dimensions 1, 3, 13, and 19 extracted from the first PC of the hyperspectral image.

The texture feature of a pixel located on  $(x, y)$  is obtained by

$$\begin{aligned} \mathbf{s}_{\text{Texture}} &= [F_{1,1}(x, y), \dots, F_{s,d}(x, y)]' \in \mathbb{R}^{60} \\ \mathbf{s}_{\text{Texture}} &= [F_{1,1}(x, y), \dots, F_{s,d}(x, y)]' \in \mathbb{R}^{60}. \end{aligned} \quad (5)$$

4) *DMP: Shape Feature*: Morphological profiles (MPs) [37], performing a series of morphological openings and closings with a family of structuring elements (SEs) of increasing size, can create an image feature set with effective discriminative ability. The opening and closing operators are basic dual morphological operators and enable us to keep the overall feature structure relatively undisturbed. In this paper, we use a reconstruction filter, which has been proved to have a better shape preservation ability than the classical morphological filters [38]. We extract the DMP feature on the first three PCs of the hyperspectral image [21].

Let  $\gamma^{\text{SE}}(I)$  and  $\phi^{\text{SE}}(I)$  be the morphological opening and closing by reconstruction with the SEs for grayscale image  $I$ .  $\mathbf{s}_{\text{MP}}$  is defined using a series of SEs with increasing sizes on  $I$

$$\begin{aligned} \mathbf{s}_{\text{MP}_\gamma} &= \left\{ s_{\text{MP}_\gamma}^\lambda(I) = \gamma^\lambda(I), \forall \lambda \in [0, n] \right\} \\ \mathbf{s}_{\text{MP}_\phi} &= \left\{ s_{\text{MP}_\phi}^\lambda(I) = \phi^\lambda(I), \forall \lambda \in [0, n] \right\} \\ &\text{with } \gamma^0(I) = \phi^0(I) = I \end{aligned} \quad (6)$$

where  $\lambda = 1, \dots, 10$  represents the radius of the disk-shaped SEs. Subsequently,  $\mathbf{s}_{\text{DMP}}$ 's are defined as vectors where the

measures of the slopes of  $\mathbf{s}_{\text{MP}}$  are stored for each step of an increasing SE series

$$\begin{aligned} \mathbf{s}_{\text{DMP}_\gamma} &= \left\{ s_{\text{DMP}_\gamma}^\lambda(I) = \left| s_{\text{MP}_\gamma}^\lambda(I_1) - s_{\text{MP}_\gamma}^{\lambda-1}(I) \right| \right. \\ &\quad \left. \forall \lambda \in [1, n] \right\} \\ \mathbf{s}_{\text{DMP}_\phi} &= \left\{ s_{\text{DMP}_\phi}^\lambda(I) = \left| s_{\text{MP}_\phi}^\lambda(I_1) - s_{\text{MP}_\phi}^{\lambda-1}(I) \right| \right. \\ &\quad \left. \forall \lambda \in [1, n] \right\}. \end{aligned} \quad (7)$$

In the proposed scheme,  $\mathbf{s}_{\text{DMP}_\gamma}$  and  $\mathbf{s}_{\text{DMP}_\phi}$  are concatenated into a  $\mathbf{s}_{\text{DMP}}$  vector in order to represent both the bright and dark features of an image. Finally, the features extracted from the first few PCs are stacked to represent the shape feature

$$\begin{aligned} \mathbf{s}_{1,\text{DMP}} &= \left\{ \mathbf{s}'_{\text{DMP}_\gamma}, \mathbf{s}'_{\text{DMP}_\phi} \right\} \\ \mathbf{s}_{\text{DMP}} &= [\mathbf{s}'_{1,\text{DMP}}, \mathbf{s}'_{2,\text{DMP}}, \mathbf{s}'_{3,\text{DMP}}]' \in \mathbb{R}^{60}. \end{aligned} \quad (8)$$

## B. CR Classification With Multitask Learning (CRC-MTL) for HSI

As mentioned previously, the spectral value and spectral gradient features reflect the characteristics of each pixel via the spectral reflectance and derivative reflectance, the Gabor texture feature can enhance the detailed edge information, and



the DMP feature focuses on the local contrast in a spatial window. In this way, each feature describes the hyperspectral scene from a single perspective and therefore cannot contain all of the discriminative information. In addition, it is widely acknowledged that to extract one optimal feature for all of the classes is not realistic. Therefore, instead of using a single feature for all classes, a better way is to combine the multiple complementary features in the image analysis [22].

Recently, based on the CR mechanism, Yang *et al.* [31] extended the traditional CR-based classification (CRC) method to the multiple feature case with a multitask learning approach (referred to as the relaxed CR model in [31]). They built the corresponding dictionary for each feature and explored the representation coefficients of the different features over their associated dictionaries for the subsequent recognition task. This approach exploits the similarity and distinctiveness of different features over the unlabeled pixel and gives a pleasing performance in face recognition [29]. In this section, we describe the corresponding classification procedure with multiple meaningful features, as applied to HSI.

We first review the CR model (CRM) [29], [30], [39] with a single feature (e.g., the spectral value feature). Supposing that we have  $M$  distinct classes, for the single feature denoted as  $k$  with  $l^k$  dimensions, we set  $\mathbf{A}_i^k \in \mathbb{R}^{l^k \times N_i}$ ,  $i = 1, \dots, M$  as the  $i$ th subdictionary whose columns are the  $N_i$  training samples from the  $i$ th class and  $N = \sum_i N_i$ . The collaborative dictionary  $\mathbf{A}^k$ , which is made up of all of the subdictionaries  $\{\mathbf{A}_i^k\}_{i=1,\dots,M}$ , maps each hyperspectral pixel into a high-dimensional space corresponding to the dictionary. The CR model of the unlabeled pixel  $s_p$  can be expressed as

$$\mathbf{s}^k = [\mathbf{A}_1^k \alpha_1^k \dots \mathbf{A}_i^k \alpha_i^k \dots \mathbf{A}_M^k \alpha_M^k] = \mathbf{A}^k \alpha^k + \varepsilon^k \quad (9)$$

where  $s^k$  represents the single feature  $k$  of the unlabeled pixel  $s_p$ ,  $\alpha_i^k$  represents the coefficient subvector over the  $i$ th subdictionary  $\mathbf{A}_i^k$ , and  $\varepsilon^k$  is the random noise. As described in [30], the CR vector contains discriminative information, to some degree, and can be used to perform the subsequent classification task. This is the motivation for representing the unlabeled pixel  $s^k$  as a weighted composition of dictionary  $\mathbf{A}^k$ . The collaborative vector  $\alpha^k$  can be easily obtained by solving the following optimization problem:

$$\hat{\alpha}^k = \arg \min_{\alpha^k} \{\|\mathbf{s}^k - \mathbf{A}^k \alpha^k\|_2^2 + \lambda \|\alpha^k\|_2^2\} \quad (10)$$

where  $\lambda$  is a positive parameter to balance the data fidelity term and the regularization term.

As for the multiple feature case, we suppose that each pixel has  $K$  different modalities of features [31]. For the unlabeled pixel  $s_p$ , we denote  $s^k$  as the  $k$ th modality of the feature to be coded,  $\mathbf{A}^k$  as the dictionary of the  $k$ th modality of the feature, and  $\alpha^k$  as the coding vector of  $s^k$  over  $\mathbf{A}^k$ . One pixel described by  $K$  features can be represented as

$$\begin{aligned} \mathbf{s}^1 &= \mathbf{A}_1^1 \alpha_1^1 + \dots + \mathbf{A}_M^1 \alpha_M^1 + \varepsilon^1 = \mathbf{A}^1 \alpha^1 + \varepsilon^1 \\ &\vdots \\ \mathbf{s}^K &= \mathbf{A}_1^K \alpha_1^K + \dots + \mathbf{A}_M^K \alpha_M^K + \varepsilon^K = \mathbf{A}^K \alpha^K + \varepsilon^K. \end{aligned} \quad (11)$$

Since the multiple different features  $\{s^k\}_{k=1,\dots,K}$  are extracted from the different perspectives of one unlabeled hyperspectral pixel and their corresponding subdictionaries  $\{\mathbf{A}^k\}_{k=1,\dots,K}$  are constructed with the features of the same training samples, it is reasonable that these features may share some similarities [31]. Therefore, it can be assumed that the representation coefficients of these features  $\{\alpha^k\}_{k=1,\dots,K}$  over their associated subdictionaries should be similar. On the other hand, since different features are distinct from each other, the representation coefficients should be diverse enough to preserve the additional complementary information, which makes the representation flexible.

In order to balance the similarity and the diversity of the different features in the linear regression, regularization is utilized to constrain the collaborative coding vector of each feature

$$\sum_{k=1}^K \omega^k \|\alpha^k - \bar{\alpha}\|_2^2 \quad (12)$$

where  $\alpha^k$ ,  $k = 1, \dots, K$ , is the coefficient vector of the  $k$ th feature vector  $s^k$  over the  $k$ th dictionary  $\mathbf{A}^k$ ,  $\bar{\alpha}$  is the mean vector of all  $\alpha^k$ , and  $\omega^k$  is the weight assigned to the  $k$ th feature. Here,  $\|\alpha^k - \bar{\alpha}\|_2^2$  is used to reduce the variance of the coefficient vector  $\alpha^k$ , which makes the representation stable, and  $\omega^k$  is used to indicate the distinctiveness of each feature vector  $s^k$ , which can make the representation flexible. This regularization can be incorporated into the multitask CR model

$$\begin{aligned} &\{[\hat{\alpha}^k, \hat{\omega}^k]\}_{k=1,\dots,K} \\ &= \arg \min_{\alpha^k, \omega^k} \sum_{k=1}^K (\|\mathbf{s}^k - \mathbf{A}^k \alpha^k\|_2^2 + \lambda \|\alpha^k\|_2^2 + \tau \omega^k \|\alpha^k - \bar{\alpha}\|_2^2) \\ &\quad \text{s.t. prior}\{\omega^k\} \end{aligned} \quad (13)$$

where  $\lambda$  and  $\tau$  are regularization parameters to trade off the data fidelity term and the two regularization terms.

Once  $\{[\hat{\alpha}^k, \hat{\omega}^k]\}_{k=1,\dots,K}$  are obtained, we can calculate the residual errors between the multiple feature signal  $\{s^k\}_{k=1,\dots,K}$  and the approximations obtained over their corresponding subdictionaries  $\{\mathbf{A}^k\}_{k=1,\dots,K}$ ,  $i = 1, \dots, M$

$$r_i = \sum_{k=1}^K \hat{\omega}^k \left\| \mathbf{s}^k - \mathbf{A}_i^k \hat{\alpha}_i^k \right\|_2^2 \quad (14)$$

where  $\hat{\alpha}_i^k$  is the subset of the coefficient vector  $\hat{\alpha}^k$  associated with class  $i$ . The label of the unlabeled pixel  $s_p$  is then determined by the minimal total residual

$$\text{class}(s_p) = \arg \min_{i=1,\dots,M} \{r_i\}. \quad (15)$$

### III. PROPOSED FRAMEWORK

In this section, we extend the multitask CR framework by incorporating the contextual neighborhood information of the image and propose a joint CR classification approach with multitask learning (JCRC-MTL) for HSI.

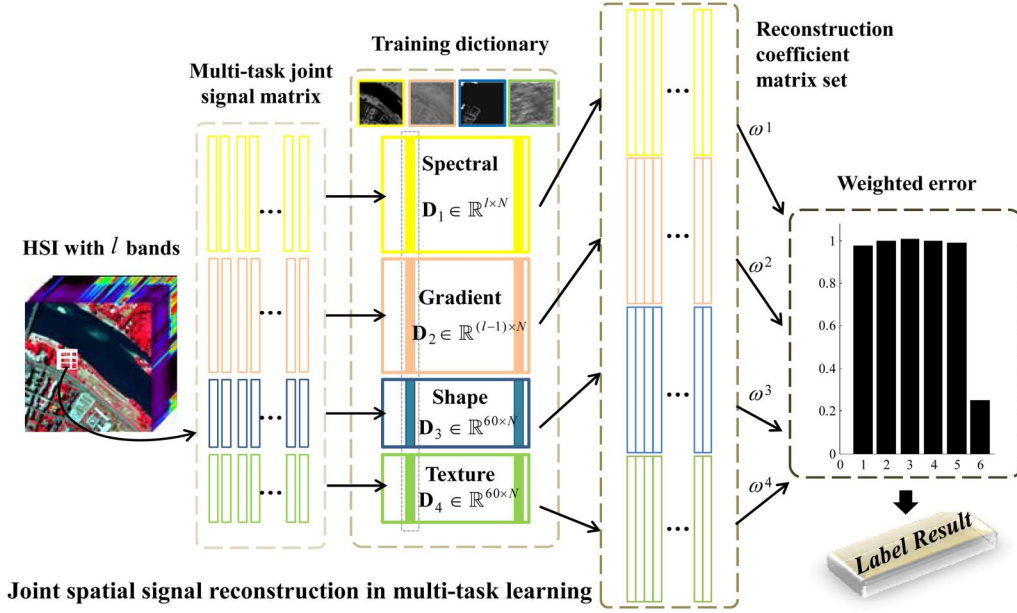


Fig. 2. Schematic illustration of the joint CR classification algorithm with multitask learning. Given a hyperspectral image, multiple modalities of the features are extracted. We construct the multitask joint signal matrices  $\{\mathbf{S}^k\}_{k=1,2,3,4}$  for the four features, and  $\mathbf{S}^k = [\mathbf{s}_1^k, \mathbf{s}_2^k, \dots, \mathbf{s}_T^k]$  with  $T$  columns, where  $T$  is the spatial neighboring window size. Each matrix is represented as a linear combination of the corresponding training feature dictionary. To preserve the diversity, the different weights of the various features are also simultaneously estimated in the linear representation procedure. Finally, the classification decision is made according to the weighted overall reconstruction error of the individual class.

#### A. Joint CR Classification Method With Multitask Learning for HSI

For the HSI case, it is usually assumed that neighboring hyperspectral pixels will consist of similar materials. Therefore, HSI pixels in a small spatial neighborhood are likely to be highly correlated and share many similarities. Previous work in [17] suggested that, by combining these contextual neighboring pixels with the unlabeled pixel into a CR in HSI, the classification accuracy can be significantly improved. For the case of only the spectral value feature, these hyperspectral pixels in a neighborhood can be simultaneously approximated by a linear combination of the common training pixels, while the common training pixels are assigned a different set of coefficients [17]. Similarly, for the multiple-feature case, these neighboring pixels can also be utilized for every feature. That is, the representation coefficients  $\{\alpha_t^k\}_{t=1,\dots,T}$  (where  $T$  is the neighborhood size) of any feature  $k$  (e.g., spectral value, spectral gradient, Gabor texture, and DMP shape feature for the hyperspectral image) in a spatial neighborhood are approximated by a collaborative linear combination of common atoms from the given dictionary  $\mathbf{A}^k$ . On the other hand, the representation coefficient set  $\{\alpha_t^k\}_{t=1,\dots,T}$  of each feature should be both stable and diverse to acquire the additional complementary information for an accurate subsequent recognition procedure. Although the multitask CR classification with these four meaningful features considers the spatial features, CRC-MTL utilizes the majority pattern of all of the contextual pixels, instead of only the unlabeled pixel, and shows a more robust performance. A visual illustration of the classification scheme for HSI with the proposed JCRC-MTL is shown in Fig. 2.

To illustrate the proposed model, consider two neighboring hyperspectral pixels  $\mathbf{s}_p$  and  $\mathbf{s}_q$ , which consist of

similar materials. With respect to the multiple feature dictionaries  $\{\mathbf{A}^k\}_{k=1,\dots,K}$ ,  $\mathbf{s}_p$  can be shown as

$$\begin{aligned} \mathbf{s}_p^1 &= \mathbf{A}_1^1 \alpha_{1,p}^1 + \dots + \mathbf{A}_M^1 \alpha_{M,p}^1 + \varepsilon_p^1 = \sum_i^M \mathbf{A}_i^1 \alpha_{i,p}^1 + \varepsilon_p^1 \\ &\vdots \\ \mathbf{s}_p^K &= \mathbf{A}_1^K \alpha_{1,p}^K + \dots + \mathbf{A}_M^K \alpha_{M,p}^K + \varepsilon_p^K = \sum_i^M \mathbf{A}_i^K \alpha_{i,p}^K + \varepsilon_p^K \end{aligned} \quad (16)$$

where  $\mathbf{A}_i^k$  denotes the  $i$ th low-dimensional subspace of the feature  $k$ ,  $\alpha_{i,p}^k$  is the corresponding coefficient subvector over the subdictionary  $\mathbf{A}_i^k$ , and  $\varepsilon_p^k$  represents the random noise associated with feature  $k$ . With the assumption that  $\mathbf{s}_p$  and  $\mathbf{s}_q$  consist of similar materials, the pixel  $\mathbf{s}_q$  can also be approximated by a linear combination of the same dominant subspace

$$\begin{aligned} \mathbf{s}_q^1 &= \mathbf{A}_1^1 \alpha_{1,q}^1 + \dots + \mathbf{A}_M^1 \alpha_{M,q}^1 + \varepsilon_q^1 = \sum_i^M \mathbf{A}_i^1 \alpha_{i,q}^1 + \varepsilon_q^1 \\ &\vdots \\ \mathbf{s}_q^K &= \mathbf{A}_1^K \alpha_{1,q}^K + \dots + \mathbf{A}_M^K \alpha_{M,q}^K + \varepsilon_q^K = \sum_i^M \mathbf{A}_i^K \alpha_{i,q}^K + \varepsilon_q^K. \end{aligned} \quad (17)$$

We next simultaneously stack all of the pixels in the neighborhood patch  $\mathbf{S} = [\mathbf{s}_1 \dots \mathbf{s}_c \dots \mathbf{s}_T]$  centered at the hyperspectral pixel  $\mathbf{s}_c$ , of size  $T$ , to construct the joint signal matrix set  $\{\mathbf{S}^k\}_{k=1,\dots,K} = \{[\mathbf{s}_1^k \mathbf{s}_2^k \dots \mathbf{s}_T^k]\}_{k=1,\dots,K}$ , which contains  $K$

matrices sized  $l^k \times T$  for each neighborhood patch. Using the joint CR model with the multitask learning procedure,  $\{\mathbf{S}^k\}_{k=1,\dots,K}$  can be represented by

$$\begin{aligned}
 \mathbf{S}^1 &= [\mathbf{s}_1^1 \dots \mathbf{s}_T^1] \\
 &= \begin{bmatrix} (\mathbf{A}_1^1 \alpha_{1,1}^1 + \dots + \mathbf{A}_M^1 \alpha_{M,1}^1 + \xi_1^1)' \\ \vdots \\ (\mathbf{A}_1^1 \alpha_{1,T}^1 + \dots + \mathbf{A}_M^1 \alpha_{M,T}^1 + \xi_T^1)' \end{bmatrix}' \\
 &= \mathbf{A}_1^1 \begin{bmatrix} (\alpha_{1,1}^1)' \\ \vdots \\ (\alpha_{1,T}^1)' \end{bmatrix}' + \dots + \mathbf{A}_M^1 \begin{bmatrix} (\alpha_{M,1}^1)' \\ \vdots \\ (\alpha_{M,T}^1)' \end{bmatrix}' + \Sigma^1 \\
 &= \sum_i^M \mathbf{A}_i^1 \Psi_i^1 + \Sigma^1 \\
 &\vdots \\
 \mathbf{S}^K &= [\mathbf{s}_1^K \dots \mathbf{s}_T^K] \\
 &= \begin{bmatrix} (\mathbf{A}_1^K \alpha_{1,1}^K + \dots + \mathbf{A}_M^K \alpha_{M,1}^K + \xi_1^K)' \\ \vdots \\ (\mathbf{A}_1^K \alpha_{1,T}^K + \dots + \mathbf{A}_M^K \alpha_{M,T}^K + \xi_T^K)' \end{bmatrix}' \\
 &= \mathbf{A}_1^K \begin{bmatrix} (\alpha_{1,1}^K)' \\ \vdots \\ (\alpha_{1,T}^K)' \end{bmatrix}' + \dots + \mathbf{A}_M^K \begin{bmatrix} (\alpha_{M,1}^K)' \\ \vdots \\ (\alpha_{M,T}^K)' \end{bmatrix}' + \Sigma^K \\
 &= \sum_i^M \mathbf{A}_i^K \Psi_i^K + \Sigma^K \quad (18)
 \end{aligned}$$

where  $\{\Psi^k\}_{k=1,\dots,K}$  is a set of the coding coefficient matrix associated with the corresponding feature dictionary  $\{\mathbf{A}^k\}_{k=1,\dots,K}$  and  $\{\Psi^k\}_{k=1,\dots,K}$  is the subset of  $\{\Psi^k\}_{k=1,\dots,K}$  over the subdictionary  $\{\mathbf{A}^k\}_{k=1,\dots,K}$ .  $\{\Sigma^k\}_{k=1,\dots,K}$  is the random noise matrix set for the neighborhood patch. In this scheme, two constraints have been included to make the label result reasonable. The first constraint assumes that the coefficients of an unlabeled pixel crossing multiple features in the proposed model should simultaneously ensure both stability and flexibility. The second prior considers that the hyperspectral pixels in a small neighboring window should share the same low-dimensional dominant subspace.

Given the training pixels, after we construct the multitask dictionaries with multiple features  $\{\mathbf{A}^k\}_{k=1,\dots,K}$ ,  $\{\Psi^k, \omega^k\}_{k=1,\dots,K}$  can be calculated by the following joint CR model with multitask learning:

$$\begin{aligned}
 &\{[\hat{\Psi}^k, \hat{\omega}^k]\}_{k=1,\dots,K} \\
 &= \arg \min_{\Psi^k, \omega^k} \sum_{k=1}^K (\|\mathbf{S}^k - \mathbf{A}^k \Psi^k\|_F^2 + \lambda \|\Psi^k\|_F^2 + \tau \omega^k \|\Psi^k - \bar{\Psi}\|_F^2) \\
 &\quad \text{s.t. prior}\{\omega^k\}. \quad (19)
 \end{aligned}$$

Assuming that the weights  $\{\omega^k\}$  can be prelearned by using a validation data set and can be fixed in advance, (19) can be denoted as

$$\{[\hat{\Psi}^k]\}_{k=1,\dots,K} = \arg \min_{\Psi^k, \omega^k} \sum_{k=1}^K (\|\mathbf{S}^k - \mathbf{A}^k \Psi^k\|_F^2 + \lambda \|\Psi^k\|_F^2 + \tau \omega^k \|\Psi^k - \bar{\Psi}\|_F^2) \quad (20)$$

where  $\bar{\Psi} = \sum_{k=1}^K \Psi^k / K$  is the mean matrix of the coefficient set  $\{\Psi^k\}_{k=1,\dots,K}$ ,  $\lambda$  and  $\tau$  are two positive parameters, and  $\omega^k$  is the weight assigned to the  $k$ th feature.

We can get a close-form solution for  $k = 1, \dots, K$

$$\Psi^k = \Psi^{0,k} + \tau \frac{\omega^k}{\sum_{\eta=1}^K \omega^\eta} \mathbf{P}^k \mathbf{Q} \sum_{\eta=1}^K \omega^\eta \Psi^{0,\eta} \quad (21)$$

$$\bar{\Psi} = \sum_{k=1}^K \omega^k \Psi^k / \sum_{k=1}^K \omega^k \quad (22)$$

where  $\mathbf{P}^k = ((\mathbf{A}^k)^T \mathbf{A}^k + \mathbf{I}(\lambda + \tau \omega^k))^{-1}$ ,  $\Psi^{0,k} = \mathbf{P}^k (\mathbf{A}^k)^T \mathbf{S}^k$ , and  $\mathbf{Q} = (\mathbf{I} - \sum_{\eta=1}^K \omega^\eta \mathbf{P}^\eta)^{-1}$  for (21), where  $\omega^\eta = \tau (\omega^\eta)^2 / \sum_{k=1}^K \omega^k$ .

For the unlabeled pixel, once the weights  $\{\hat{\omega}^k\}_{k=1,\dots,K}$  and the coefficient matrix set  $\{\hat{\Psi}^k\}_{k=1,\dots,K}$  for all of the features are obtained, the overall coding error for class  $i$  is shown as

$$r_i = \sum_{k=1}^K \hat{\omega}^k \|\mathbf{S}^k - \mathbf{A}_i^k \hat{\Psi}_i^k\|_F^2 \quad (23)$$

where  $\hat{\Psi}_i^k$  is the coefficient matrix associated with feature  $k$  and class  $i$ . The label of the unlabeled pixel  $\mathbf{s}_c$  is then determined by the minimal total residual

$$\text{class}(\mathbf{s}_c) = \arg \min_{i=1,\dots,M} \{r_i\}. \quad (24)$$

The computational burden for the proposed JCRC-MTL algorithm can be shown as follows. For the  $k$ th feature, we suppose that the size of  $\mathbf{S}^k$  is  $l^k \times T$  and the size of dictionary  $\mathbf{A}^k$  is  $l^k \times n$  (where  $l^k$  is the dimensionality of the  $k$ th feature,  $T$  denotes the size of the spatial window, and  $n$  refers to the number of training samples of the data set). When the weight values are known, the time complexity of the closed-form solution (21) is  $O(\sum_{k=1}^K (2n^2 + n^2 T + n T l^k))$ , as the  $\mathbf{P}^k (\mathbf{A}^k)^T \mathbf{S}^k$  costs  $O(n^2 + n T l^k)$  and the  $\mathbf{P}^k \mathbf{Q} \sum_{\eta=1}^K \omega^\eta \Psi^{0,\eta}$  costs  $O(n^2 + n^2 T)$ .

## B. Final Classification Scheme of HSI

The implementation details of the proposed JCRC-MTL algorithm for HSI are shown in **Algorithm 1**.

**Algorithm 1:** The JCRC-MTL algorithm for HSI**Input:**

- 1) A HSI, in which each pixel  $s_c$  is represented as  $\{s_c^k\}_{k=1,\dots,K}$
- 2) Regularization parameters  $\lambda$  and  $\tau$ , spatial neighborhood size  $T$ , and the given weight vector for every feature  $\omega^k = \{\omega^k\}_{k=1,\dots,K}$

**Initialization:**

- 1) Multiple feature extraction from the hyperspectral image
- 2) Construct the multiple features dictionaries  $\{\mathbf{A}^k\}_{k=1,\dots,K}$  with all of the training set in this hyperspectral image, and normalize each column of  $\mathbf{A}^k$  to have unit  $\ell_2$ -norm

**For** each unlabeled pixel in the hyperspectral image:

- 1) Construct the multiple feature joint collaborative matrix set  $\{\mathbf{S}^k\}_{k=1,\dots,K}$
- 2) Code  $\{\mathbf{S}^k\}_{k=1,\dots,K}$  over the dictionaries  $\{\mathbf{A}^k\}_{k=1,\dots,K}$ , and acquire the coefficient matrices  $\{\Psi^k\}_{k=1,\dots,K}$  via (21) and (22)
- 3) Compute the regularized residuals and label the unlabeled pixel via (23) and (24)
- 4) Turn to the next pixel

**End For****Output:** A 2-D matrix which records the labels of the HSI

## IV. EXPERIMENTS

In this section, we utilize two classical hyperspectral data sets to validate the effectiveness and superiority of the proposed JCRC-MTL algorithm. All of the experiments were carried out using MATLAB on a PC with one 3.50-GHz processor and 16.0 GB of RAM.

*A. ROSIS Urban Data Set: Pavia University, Italy*

1) *Data Set Description:* This scene was acquired by the ROSIS sensor during a flight campaign over Pavia University, northern Italy. The number of available spectral bands is 103, and the geometric resolution is 1.3 m. The image size is  $610 \times 610$  pixels, and we cut a patch sized  $610 \times 340$ . The false color composite of the Pavia University image is shown in Fig. 4(a). Based on the feature extraction procedure described in Section II-A, we obtain the spectral value vector  $s_{\text{Spectral}} \in \mathbb{R}^{103}$ , the spectral gradient vector  $s_{\text{Gradient}} \in \mathbb{R}^{102}$ , the Gabor texture vector  $s_{\text{Gabor}} \in \mathbb{R}^{60}$ , and the DMP shape feature  $s_{\text{DMP}} \in \mathbb{R}^{60}$ , in which radius  $\lambda_{\text{DMP}} = \{1, 2, \dots, 10\}$  and  $b = \{\text{PC1}, \text{PC2}, \text{PC3}\}$  for each pixel in the HSI. This image contains nine ground-truth classes, as shown in Table I. The ground truth is visually shown in Fig. 4(b).

2) *Evaluation of the Multiple Feature Combination:* This section is utilized to demonstrate the superiority of the multiple feature combination capability in the proposed JCRC-MTL algorithm. From the ground truth, we randomly select 30 pixels

TABLE I  
NINE GROUND-TRUTH CLASSES OF THE ROSIS  
PAVIA UNIVERSITY DATA SET

Class name	Samples	Class name	Samples
Asphalt	6631	Bare soil	5029
Meadows	18649	Bitumen	1330
Gravel	2099	Brick	3682
Trees	3064	Shadows	947
Metal sheet	1345	Total	42756

for each class as the training samples and the rest as the test samples to validate the performances. The following classifiers are used for the comparisons: the CR classifier with each single feature, the SVM classifier with each single feature, and the SVM classifier which directly stacks all four feature vectors into an augmented higher dimensional feature [21] (referred to as SVM-VS in this paper). Every SVM-based classifier in this experiment uses the radial basis function (RBF) kernel [12], [13]. The CR classification with multiple-task learning [31] (CRC-MTL), which can be considered as a special case of JCRC-MTL when the joint feature signal set is just the feature vector set of the central unlabeled pixel itself, is also included in the comparisons. All of the parameters for every classifier in this experiment are obtained by tenfold cross-validation (CV) [40]. For the multitask-learning-based algorithms, we fix each feature equally and set  $\omega^k = 0.25$ . The classification accuracies are averaged over ten runs for each classifier to reduce the possible bias induced by the random sampling.

The classification accuracy for each class, the average OA with the associated standard deviation, and the average kappa coefficient with the standard deviation, using the different classifiers, are shown in Table II for the test data. The  $z$  value [41], which reflects the significant differences between two classifiers, is also included in Table II. The  $z$  test value can be computed by  $z = \hat{\kappa}_1 - \hat{\kappa}_2 / \sqrt{\hat{\sigma}_{\kappa_1}^2 + \hat{\sigma}_{\kappa_2}^2}$ , where  $\hat{\kappa}_1$  and  $\hat{\kappa}_2$  are the estimated kappa coefficient values for the related classifier and  $\hat{\sigma}_{\kappa_1}^2$  and  $\hat{\sigma}_{\kappa_2}^2$  represent the estimated variances of the derived kappa coefficients. For the widely used 5% level of significance, we consider that there is a significant difference between the two classifications if  $|z| > 1.96$  holds. In this part, we take the classification result of the proposed JCRC-MTL as the baseline and calculate the  $z$  values of the other classification results. Each value in the last row of Table II represents the average running time of the associated classifier. In Table II, the best results for each quality index are labeled in bold, and the optimal results for each single-feature-based classifier are underlined. Due to space constraint, we have omitted the visual classification maps in this section, while some visual maps of these 11 classifiers are shown in Fig. 4. Overall, it can be seen that the proposed JCRC-MTL algorithm achieves the best performance for each class.

We now investigate the complementary properties of the aforementioned multiple features. Fig. 3 shows the correlation coefficient matrix across the different features for this data set. The first 103 dimensions correspond to the spectral value feature, the next 102 dimensions for the spectral gradient feature, the following 60 dimensions for the Gabor texture feature, and the last 60 dimensions for the DMP shape feature. It can be observed that the intrafeature correlations are generally smaller



TABLE II  
CLASSIFICATION ACCURACY FOR THE PAVIA UNIVERSITY DATA SET WITH THE TEST  
SET IN THE EVALUATION OF THE MULTIPLE FEATURE COMBINATION SECTION

Method	CRC				SVM					CRC-M TL	JCRC- MTL
Features	Spectral value	Spectral gradient	Gabor texture	DMP shape	Spectral value	Spectral gradient	Gabor texture	DMP shape	SVM- VS	MTL with four features	
1	0.3095	0.3256	0.5461	<u>0.8334</u>	0.6803	0.4809	0.4251	<u>0.8752</u>	0.9773	0.9135	<b>0.9766</b>
2	0.7168	0.5940	<u>0.8066</u>	0.5872	0.7106	0.6187	0.6587	<u>0.7278</u>	0.8494	0.8964	<b>0.9455</b>
3	0.8403	0.5880	0.8776	<u>0.8944</u>	0.7360	0.6732	0.5715	<u>0.9365</u>	0.9352	0.9564	<b>0.9847</b>
4	0.9467	0.9597	0.3952	<u>0.9638</u>	0.9097	0.9443	0.3860	<u>0.9463</u>	0.9613	0.9905	<b>0.9752</b>
5	<u>0.9992</u>	0.9984	0.8373	0.9965	0.9921	0.9957	0.8140	<u>0.9976</u>	0.9944	0.9993	<b>1</b>
6	0.5415	0.4838	0.7417	<u>0.8739</u>	0.7401	0.5939	0.6245	<u>0.8741</u>	0.9145	0.9421	<b>0.9905</b>
7	0.7626	0.3915	0.7690	<u>0.9614</u>	0.8848	0.5677	0.6085	<u>0.9455</u>	0.9506	0.9735	<b>0.9978</b>
8	0.1460	0.2964	0.6783	<u>0.9268</u>	0.7611	0.4809	0.5601	<u>0.9316</u>	0.9180	0.9679	<b>0.9877</b>
9	0.9534	0.6900	0.6184	<u>0.9755</u>	<u>0.9819</u>	0.8963	0.6659	0.9750	0.9818	0.9942	<b>0.9855</b>
OA	0.6216± 0.0220	0.5480± 0.0184	0.7076± 0.0208	0.7546± 0.0347	0.7490± 0.0319	0.6248± 0.0206	0.5896± 0.0678	0.8315± 0.0209	0.9054± 0.0210	0.9279± 0.0194	<b>0.9674± 0.0128</b>
$\kappa$	0.5236± 0.0254	0.4452± 0.0165	0.6233± 0.0235	0.7005± 0.0385	0.6826± 0.0342	0.5338± 0.0225	0.4927± 0.0727	0.7853± 0.0252	0.8776± 0.0264	0.9063± 0.0245	<b>0.9573± 0.0166</b>
$z$	-14.298	-21.903	-11.613	-6.125	-7.229	-15.144	-6.2317	-5.705	-2.559	-0.9956	<b>0</b>
Time	6.1	6.6	<b>4.8</b>	5.6	23.9	31.5	34.5	19.8	122.2	674.7	738.4

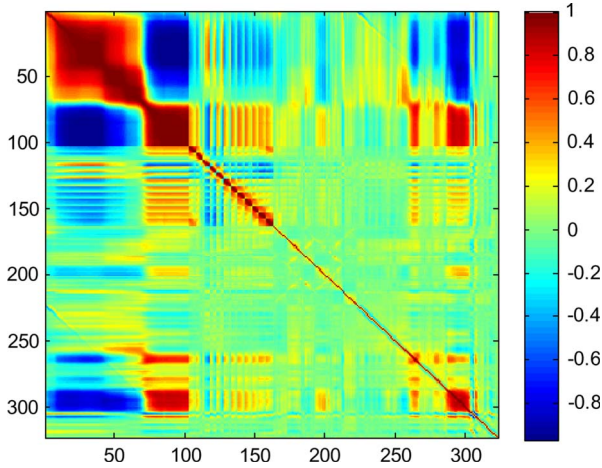


Fig. 3. Correlation coefficient matrix across different features for the Pavia University data set.

than the interfeature correlations, which preliminarily suggests the complementary properties of the four features. The two classifiers, CRC and SVM, are implemented here to reduce the possible bias induced by a single classification algorithm. Comparing these classification results in the first eight columns in Table IV, it is shown that almost all of the features, except for the spectral gradient feature, can achieve the best performance on certain classes. For instance, with the CRC method, the spectral value feature achieves the best classification result with class 5, the Gabor texture feature obtains the best result with class 2, and the DMP shape feature performs the best on the rest of the classes. Similar observations can be made for the SVM classification method. That is to say, different features can reflect different aspects of the discriminative information of the HSI and are complementary to each other. It is therefore natural to combine the four features to improve the classification result.

This point is validated by the classification results of SVM-VS and CRC-MTL. Here, it can be seen that the SVM-VS approach has an improvement of 0.0739 on OA over the best result of the SVM classifier with a single feature, and the CRC-MTL method shows a remarkable improvement of 0.1733 over the CRC method with a single feature. It is also demonstrated that the improvement of CRC-MTL over CRC is greater than that of the SVM-VS approach over SVM with a single feature, which validates the superiority of the multitask learning approach. Finally, we focus on the effect of the joint spatial signals in the JCRC-MTL method. Here, it is also shown that, with the help of the neighboring pixels, the classification accuracy can be further improved, reaching an OA of 0.9674. In conclusion, it is demonstrated that the discrimination of these features between different classes can be further enhanced, with more robustness, in the proposed JCRC-MTL method.

3) *Classification Result*: In this section, we further analyze the accuracy of the proposed JCRC-MTL algorithm with the Pavia University data set. The aforementioned ten random independent classifications with 30 training samples per class are utilized in this section. The single-feature-based algorithms consist of SVM with the RBF kernel, CRC, SRC with an improved  $\ell_1$ -norm algorithm called lasso [42], [43], and JSRC [17] with the greedy pursuit algorithm (referred to as SOMP in [44]). As suggested in the last section, since the spectral value feature and the DMP-feature-based classifiers lead to better accuracies than the others, each single-feature-based classification algorithm has the two instances. For example, CRC-S refers to the classifier utilizing the spectral value feature via the CRC classification, and CRC-D refers to the classifier utilizing the DMP feature. For the multiple-feature-based algorithm, we also first utilize the SVM-VS, with regard to its desirable performance. In addition, two classifiers with multitask learning are also included in the comparisons. One is CRC-MTL, and the



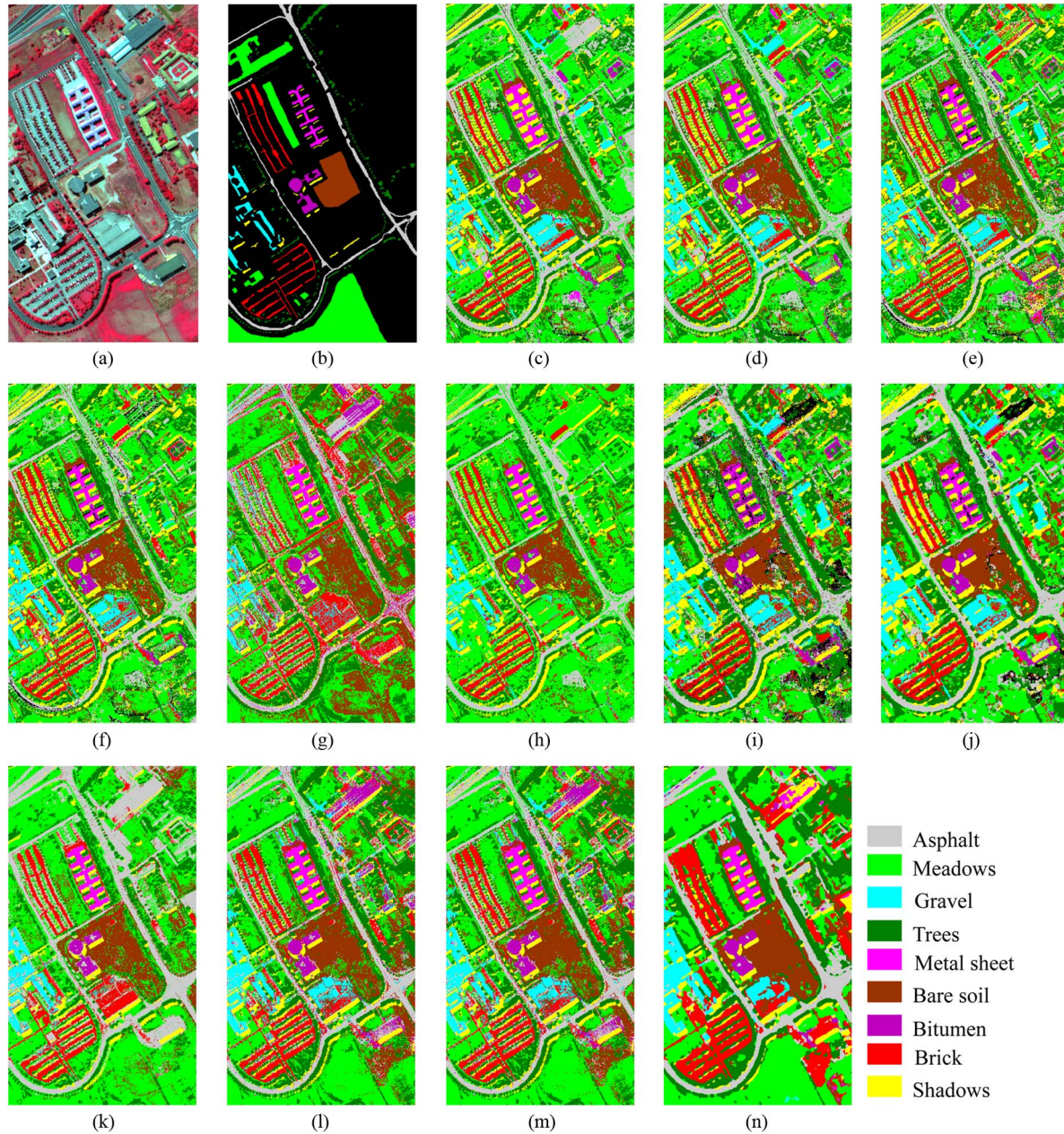


Fig. 4. Classification results with the Pavia University image. (a) False color image (R:102, G:56, and B:31). (b) Ground truth. (c) CRC-S. (d) CRC-D. (e) SRC-S. (f) SRC-D. (g) SVM-S. (h) SVM-D. (i) JSRC-S. (j) JSRC-D. (k) SVM-VS. (l) SRC-MTL. (m) CRC-MTL. (n) JCRC-MTL.

other is the sparse representation classification with multitask learning (SRC-MTL) algorithm (also referred to as MTJSRC in [26]).

The detailed analysis of the classification result for the Pavia University data set is described in this section. The range of the parameters for CV can be shown as follows: the regularization parameters  $\lambda$  and  $\tau$  for the three multitask-learning-based classification algorithms range from  $1e-6$  to  $1e-1$ , and the neighborhood size for JCRC-MTL and JSRC ranges from 1 to 169. Parameter settings for the other classifiers are also selected by tenfold CV from a reasonable range.

The thematic maps of the various classifiers are visually shown in Fig. 4(c)–(n), respectively. For the test data, the quantitative results for each class, the average OA with the

associated standard deviation, and the average kappa coefficient with the standard deviation using the different classifiers are shown in Table III. Comparing the classification results of CRC-MTL, SRC-MTL, and SVM-VS with those of the single-feature-based classifiers, it can be seen that the multiple features can indeed offer additional complementary information for the classification, and they improve the classification result over that of a single feature. It is also shown that the multiple-feature-based classifiers require much more computational burden than the others, while the CRC-based classifier is the fastest among the single-feature-based classifiers. A detailed running time analysis for the four multiple-feature-based classifiers is provided in Section IV-A5. In light of the recognition rate, the improvements of JSRC over SRC suggest that the spatial

TABLE III  
CLASSIFICATION ACCURACY FOR THE PAVIA UNIVERSITY IMAGE WITH THE TEST SET

C	CRC		SRC		SVM		JSRC		SVM-	SRC-M	CRC-	JCRC-
	S	D	S	D	S	D	S	D	VS	TL	MTL	MTL
1	0.3095	0.8334	0.6849	0.8489	0.6803	0.8752	0.5001	0.9103	0.9773	0.8948	0.9135	<b>0.9766</b>
2	0.7168	0.5872	0.7022	0.6129	0.7106	0.7278	0.8845	0.6117	0.8494	0.8904	0.8964	<b>0.9455</b>
3	0.8403	0.8944	0.7885	0.9469	0.7360	0.9365	0.8430	0.9743	0.9352	0.9563	0.9564	<b>0.9847</b>
4	0.9467	0.9638	0.9381	0.9539	0.9097	0.9463	0.9137	0.9482	0.9613	0.9880	<b>0.9905</b>	0.9752
5	0.9992	0.9965	0.9978	0.9977	0.9921	0.9976	<b>1</b>	0.9996	0.9944	0.9997	0.9993	<b>1</b>
6	0.5415	0.8739	0.7751	0.8336	0.7401	0.8741	0.6889	0.8843	0.9145	0.9324	0.9421	<b>0.9905</b>
7	0.7626	0.9614	0.8685	0.9596	0.8848	0.9455	0.9368	0.9591	0.9506	0.9697	0.9735	<b>0.9978</b>
8	0.1460	0.9268	0.5600	0.9363	0.7611	0.9316	0.4531	0.9449	0.9180	0.9624	0.9679	<b>0.9877</b>
9	0.9534	0.9755	0.9700	0.9826	0.9819	0.9750	0.9473	0.9894	0.9818	0.9914	<b>0.9942</b>	0.9855
OA	0.6216±	0.7546±	0.7546±	0.7619±	0.7490±	0.8315±	0.7713±	0.7666±	0.9054±	0.9150±	0.9279±	<b>0.9674±</b>
	0.0220	0.0347	0.0347	0.0276	0.0319	0.0209	0.0180	0.0528	0.0210	0.0155	0.0194	<b>0.0128</b>
$\kappa$	0.5236±	0.7005±	0.7005±	0.7079±	0.6826±	0.7853±	0.7008±	0.7162±	0.8776±	0.8898±	0.9063±	<b>0.9573±</b>
	0.0254	0.0385	0.0385	0.0298	0.0342	0.0252	0.0225	0.0575	0.0264	0.0194	0.0245	<b>0.0166</b>
$z$	-14.298	-6.125	-27.491	-27.303	-7.229	-5.705	-9.171	-3.185	-2.559	-1.659	-0.9956	<b>0</b>
Time	6.1	<b>5.6</b>	18.0	13.0	23.9	19.8	93.4	22.0	122.2	173.1	674.7	738.4

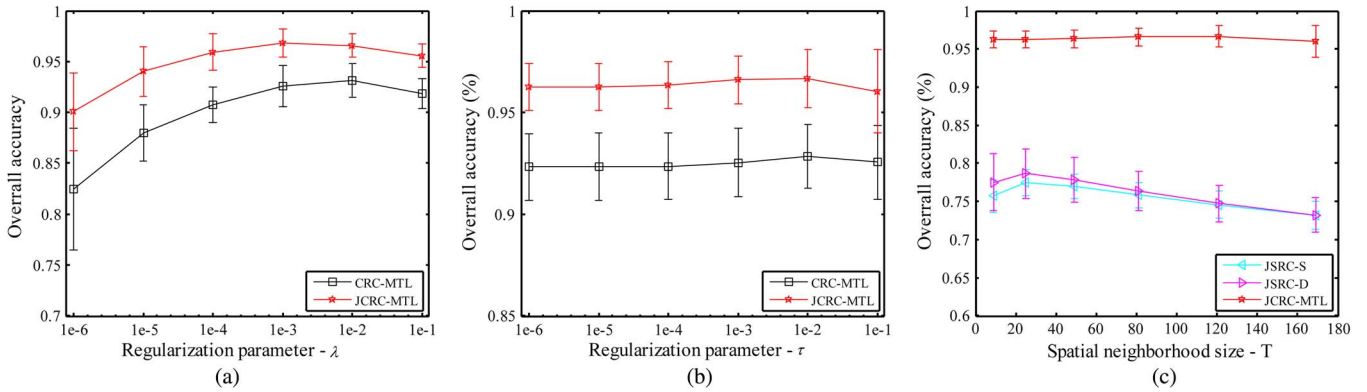


Fig. 5. Classification accuracy versus parameters for the various classification algorithms. (a) Regularization parameter  $\lambda$  for JCRC-MTL and CRC-MTL. (b) Regularization parameter  $\tau$  for JCRC-MTL and CRC-AWMTL. (c) Size of the spatial neighborhood window  $T$  for JSRC-S, JSRC-D, and JCRC-MTL.

neighborhood pixels can also improve the classification effectiveness. By simultaneously incorporating the multiple features and the spatial neighborhood information, the proposed JCRC-MTL method achieves the best results for most classes and the best OA and kappa coefficient and shows a significant improvement over the classical SVM method.

4) *Parameter Analysis*: In this section, we examine the effect of the parameters on the classification performance of the various algorithms with the Pavia University image. We fix the other parameters as the corresponding optimal and focus on one specific parameter. The experiments for  $\lambda$ ,  $\tau$ , and  $T$  are repeated ten times using different randomly chosen training sets to reduce the possible bias induced by the random sampling. The horizontal axis shown in each subfigure in Fig. 5 is the value range of the corresponding parameter, while the vertical axis shows the OA (in percent) of the different classifiers.

In Fig. 5(a), it can be seen that the performance of all of the multitask-learning-based classification algorithms generally improves as the regularization parameter  $\lambda$  increases and then begins to decrease after the maximum value, and JCRC-MTL

shows a better and more stable performance than CRC-MTL. Since  $\lambda$  makes a tradeoff between the data fidelity term and the prior term of the coefficient matrix for each feature, it is demonstrated that the multitask-based algorithms are quite robust to this parameter over a wide range of values.

In Fig. 5(b), both JCRC-MTL and CRC-MTL are quite stable and achieve a pleasing accuracy when the regularization parameter  $\tau$  ranges from  $1e-6$  to  $1e-1$ . Once  $\tau$  exceeds a certain threshold (denoted as  $1e-2$  in this experiment), the dominant part of (13) becomes the last term  $\omega^k \|\Psi^k - \tilde{\Psi}\|_F^2$ , which will weaken the discriminative power for the subsequent classification.

Finally, we range  $T$  from 9 to 169 to investigate the effect of the neighborhood size. In Fig. 5(c), it can be seen that the classification result of the proposed method is slightly better than those of the JSRC-based classifiers for all of the neighborhood sizes. After reaching the extreme point, the accuracy curve of the proposed method is more stable than that of the other two methods. It can therefore be concluded that JCRC-MTL is robust with regard to different neighborhood sizes.



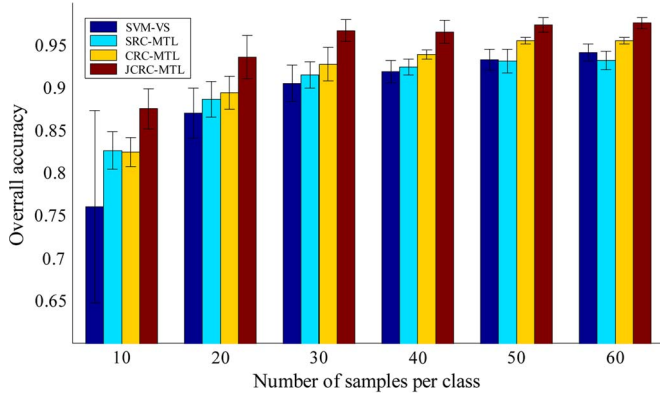


Fig. 6. Effect of the number of training samples per class  $N_i$  for the Pavia University data set. For fairness, the number of training pixels for each class is equal.

TABLE IV  
RUNNING TIME ANALYSIS FOR THE FOUR MULTIPLE-FEATURE-BASED CLASSIFIERS FOR THE PAVIA UNIVERSITY DATA SET

	10	20	30	40	50	60
SVM-VS	49.5	71.1	122.2	133.3	172.5	217.6
SRC-MTL	51.1	67.6	111.3	141.3	214.1	296.4
CRC-MTL	125.2	243.1	674.7	1435.8	2427.7	3950.2
JCRC-MTL	135.0	367.4	738.4	1594.0	2464.0	3997.2

5) *Running Time and the Effect of the Number of Training Samples:* Next, we show the effect of the number of training samples on the classification accuracy of the four multiple-feature-based algorithms, as shown in Fig. 6. For fairness, the number of training pixels for each class is equal, and we randomly choose 10–60 labeled pixels in each class as the training samples and the remainder as the test ones. The horizontal axis in Fig. 6 is the number of training samples, while the vertical axis shows the OA of the different algorithms. The accuracy is averaged over ten runs at each point to reduce the possible bias induced by the random sampling. It is immediately clear from Fig. 6 that, in most cases, the overall accuracy increases monotonically as the training samples increase and so does the stability of each classifier. For the small sample set case, SVM-VS shows an inferior performance to the others, which suggests that the multitask-learning-based algorithms are more robust, especially the proposed JCRC-MTL. With the increase of the training samples, SVM-VS gradually tends to be comparable to SRC-MTL. Compared with the first three classifiers, it is demonstrated that the CR is more suitable than a sparse representation, and the multitask learning enables a more elegant performance than the well-used vector stacking approach when dealing with multiple-feature-related cases. It is further shown in Fig. 6 that the final proposed JCRC-MTL can give a state-of-the-art result even when the number of training samples is quite limited, which further verifies the superiority of the joint contextual information imposed into the proposed algorithm.

For the running time comparison, the detailed average running times for each case of each classifier are shown in Table IV. It is shown that the first two classifiers have comparable running times, and they are faster than the latter CR-based classifiers. Although the running time of SVM-VS seems to

TABLE V  
NINE GROUND-TRUTH CLASSES OF THE AVIRIS INDIAN PINES IMAGE DATA SET

Class name	Samples	Class name	Samples
Corn-notill	1428	Soybean-notill	972
Corn-mintill	830	Soybean-mintill	2455
Grass-pasture	483	Soybean-clean	593
Grass-trees	730	Woods	1265
Hay-windrowed	478	Bldg-grass-trees	386
Total			9620

be optimal, it should be noted that the SVM is implemented by the LibSVM [45] package, which utilizes C++ software to speed it up. It should also be noted that the speed superiority of SRC-MTL over CRC-MTL is mainly caused by the accelerated proximal gradient [26] method utilized in the algorithm, which accelerates the convergence speed. When dealing with a large-scale training sample set as the dictionary, CRC-MTL and JCRC-MTL, utilizing the matrix inverse operation, will be left behind, as shown in Table IV. In view of this, it is considered that SRC-MTL is the fastest, and the proposed algorithm requires more computing time, due to the CR with multitask learning and the extended contextual information 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 soon no longer be an issue.

## B. 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 it consists of  $145 \times 145$  pixels and 220 spectral reflectance bands in the wavelength range of 0.4–2.5  $\mu\text{m}$ . The false color composite of the Indian Pines image is shown in Fig. 7(a). We reduced the number of bands to 200 by removing the bands covering the regions of water absorption: 104–108, 150–163, and 220, as in [46]. The spatial resolution for this image is about 20 m. The feature extraction procedure is the same as that utilized in the former experiment. From the reference data, we randomly sample 30 pixels for each class as the training samples and the rest as the test set. This image contains ten ground-truth classes, and details of the quantities are shown in Table V. In addition, the ground truth is visually shown in Fig. 7(b).

The thematic results using CRC, SRC, SVM with the RBF kernel, JSRC, CRC-MTL, MTJL-SRC, and JCRC-MTL are visually shown in Fig. 7(c)–(n), respectively. The quantitative evaluations, which include the classification accuracy for each class, the OA, the kappa coefficient value  $\kappa$ , the running time, and the  $z$  test value, which all take the result of JCRC-MTL as the baseline, are shown in Table VI. The experimental design of this table is the same as that of Table III, and we again give the average results over ten runs for each classifier. The ranges of the parameters are the same as those for the ROSIS image experiment, and all of the parameters for each classifier are obtained by tenfold CV. Overall, it can be concluded that the proposed JCRC-MTL method yields the best performance by simultaneously incorporating the multiple features and the neighborhood information.



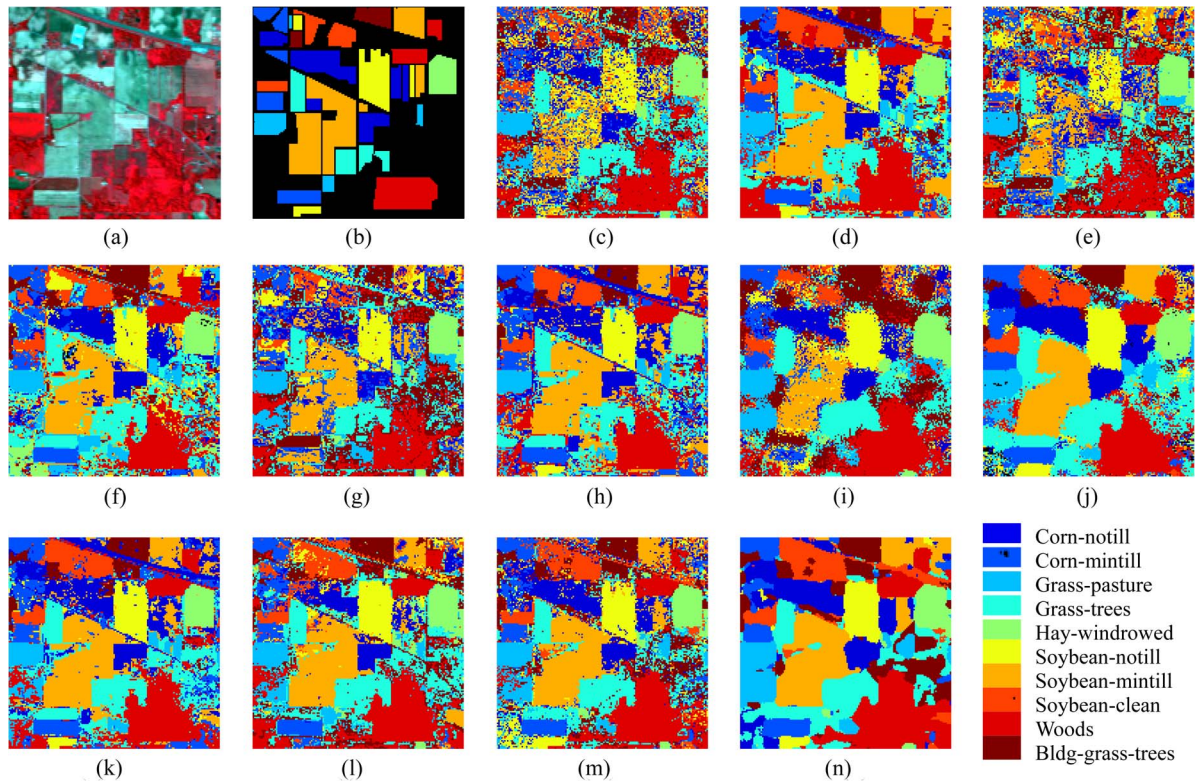


Fig. 7. Classification results with the Indian Pines image. (a) False color image (R:57, G:27, and B:17). (b) Ground truth. (c) CRC-S. (d) CRC-D. (e) SRC-S. (f) SRC-D. (g) SVM-S. (h) SVM-D. (i) JSRC-S. (j) JSRC-D. (k) SVM-VS. (l) SRC-MTL. (m) CRC-MTL. (n) JCRC-MTL.

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

C	CRC		SRC		SVM		JSRC		SVM-VS	SRC-M	CRC-MTL	JCRC-MTL
	S	D	S	D	S	D	S	D	VS	TL	MTL	MTL
1	0.5506	0.7996	0.4535	0.8386	0.6069	0.8151	0.7000	0.8881	0.8379	0.6931	0.7770	<b>0.8532</b>
2	0.4623	0.9225	0.4931	0.9126	0.6392	0.9158	0.7546	0.9417	<b>0.9438</b>	0.8601	0.9100	0.9434
3	0.8163	0.8784	0.8483	0.8914	0.8967	0.8667	0.8698	0.9146	0.8929	0.8711	0.8996	<b>0.9174</b>
4	0.9123	0.9181	0.9170	0.9210	0.9321	0.9137	0.9836	0.9720	0.9687	0.9887	0.9867	<b>0.9963</b>
5	0.9891	0.9895	0.9929	0.9940	0.9951	0.9942	0.9982	0.9979	0.9964	0.9998	0.9998	<b>1</b>
6	0.5758	0.7904	0.6438	0.8460	0.6820	0.8200	0.7530	<b>0.8822</b>	0.8451	0.8173	0.8297	0.8581
7	0.4066	0.8475	0.4827	0.8183	0.5455	0.8200	0.5724	0.8578	0.8597	0.8294	0.8494	<b>0.8936</b>
8	0.5826	0.9142	0.5250	0.9015	0.6357	0.8961	0.6845	0.9418	0.8998	0.9188	0.9094	<b>0.9780</b>
9	0.8290	0.9662	0.8228	0.9529	0.8470	0.9525	0.8871	0.9929	0.9760	0.9572	0.9861	<b>0.9987</b>
10	0.6287	0.9590	0.5899	0.9581	0.6388	0.9584	0.8520	0.9857	0.9719	0.9359	0.9677	<b>0.9972</b>
OA	0.6110±	0.8786±	0.6221±	0.8784±	0.6932±	0.8726±	0.7504±	0.9134±	0.9007±	0.8341±	0.8880±	<b>0.9252±</b>
	0.0221	0.0212	0.0146	0.0228	0.0122	0.0227	0.0198	0.0195	0.0188	0.0196	0.0176	<b>0.0072</b>
$\kappa$	0.5577±	0.8597±	0.5689±	0.8596±	0.6485±	0.8527±	0.7156±	0.8999±	0.8850±	0.8097±	0.8704±	<b>0.9133±</b>
	0.0243	0.0241	0.0166	0.0259	0.0137	0.0257	0.0220	0.0222	0.0215	0.0220	0.0202	<b>0.0082</b>
$z$	-13.887	-2.108	-13.450	-2.099	-16.611	-2.246	-8.419	-0.568	-1.234	-4.418	-1.971	<b>0</b>
Time	1.9	<b>1.3</b>	5.2	22.6	17	11.4	46.3	12.3	65	144.0	197.9	206.5

We now analyze the performances of all of the classifiers in detail. For the single-feature-based classifiers, it can be clearly observed that the DMP-feature-based classifiers are far superior to the spectral-feature-based ones, which suggests the superiority of the MP for preserving the discriminative information between different classes. Most of the multiple-feature-based classifiers, except for SRC-MTL, show superior classification results to the counterpart single-feature-based ones, which indicates that the multiple features offer additional

useful information to improve the discrimination, even though this image is not a high spatial resolution image. For the multiple-feature-based classifiers, it is first observed that the inferior accuracy of SRC-MTL, which is even weaker than the direct feature stacking approach (referred to as SVM-VS), suggests the limitations of this algorithm in this experiment. Comparing the two  $z$  values from JCRC-MTL and CRC-MTL, it is noted that the incorporation of the neighborhood information in the proposed algorithm leads to a significant

improvement, with regard to the spatial smoothness of the image. Similar observations can be made in the comparison between JSRC and SRC.

## V. CONCLUSION

This paper has proposed a novel joint CR classification method with multitask learning (JCRC-MTL) for HSI. In the proposed approach, an HSI pixel is depicted with multiple complementary features from different perspectives, which are simultaneously integrated in the CR-based classification framework via the multitask learning approach. The neighborhood information is utilized to further improve the performance of the proposed JCRC-MTL classification method. The proposed JCRC-MTL classification method was tested on ROSIS and AVIRIS hyperspectral images, and the extensive experimental results confirm the effectiveness of the proposed hyperspectral classifier.

However, the proposed multitask learning framework could still be further improved in certain aspects. For instance, how to select the more-correlated neighboring pixels in the neighborhood patch to the test one has not been considered in the current algorithm. Therefore, our future work will focus on how to construct a more meaningful signal matrix for each feature.

## ACKNOWLEDGMENT

The authors would like to thank Prof. D. Landgrebe from Purdue University, West Lafayette, IN, USA, for providing the AVIRIS image of Indian Pines, Prof. Gamba from the University of Pavia, Pavia, Italy, for providing the ROSIS data set, and the handling editor and anonymous reviewers for their careful reading and helpful remarks.

## REFERENCES

- [1] H. Zhang, L. Zhang, and H. Shen, "A super-resolution reconstruction algorithm for hyperspectral images," *Signal Process.*, vol. 92, no. 9, pp. 2082–2096, Sep. 2012.
- [2] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, Mar. 2013.
- [3] L. Genc, M. Inalpulat, U. Kizil, M. Mirik, S. E. Smith, and M. Mendes, "Determination of water stress with spectral reflectance on sweet corn (*Zea mays* L.) using classification tree (CT) analysis," *Žemdirbystė (Agriculture)*, vol. 100, no. 1, pp. 81–90, 2013.
- [4] 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.
- [5] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.
- [6] C.-I. Chang, *Hyperspectral Data Exploitation: Theory and Applications*. Hoboken, NJ, USA: Wiley, 2007.
- [7] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. 5th Annu. Workshop Comput. Learn. Theory*, 1992, pp. 144–152.
- [8] V. Vapnik, *The Nature of Statistical Learning Theory*. Berlin, Germany: Springer-Verlag, 1999.
- [9] 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, no. 1, pp. 363–392, 2005.
- [10] 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 Recog.*, 2006, pp. 2126–2136.
- [11] L. Bruzzone, M. Chi, 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.
- [12] 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.
- [13] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, Nov. 2008.
- [14] M. Pal and G. M. Foody, "Feature selection for classification of hyperspectral data by SVM," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2297–2307, May 2010.
- [15] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson, "SVM- and MRF-based method for accurate classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 736–740, Oct. 2010.
- [16] L. Zhang, M. Yang, X. Feng, Y. Ma, and D. Zhang, "Collaborative representation based on classification for face recognition," arXiv preprint arXiv:1204.2358.
- [17] 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.
- [18] 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. 2013.
- [19] H. Zhang, J. Li, Y. Huang, and L. Zhang, "A nonlocal weighted joint sparse representation classification method for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 10.1109/JS-TARS.2013.2264720, to be published.
- [20] X. Huang and L. Zhang, "An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 12, pp. 4173–4185, Dec. 2008.
- [21] X. Huang and L. Zhang, "An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 257–272, Jan. 2013.
- [22] P. Gehler and S. Nowozin, "On feature combination for multiclass object classification," in *Proc. IEEE 12th Int. Conf. Comput. Vision*, 2009, pp. 221–228.
- [23] L. Zhang, L. Zhang, D. Tao, and X. Huang, "On combining multiple features for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 879–893, Mar. 2012.
- [24] L. O. Jimenez-Rodriguez, E. Arzuaga-Cruz, and M. Vélez-Reyes, "Unsupervised linear feature-extraction methods and their effects in the classification of high-dimensional data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 2, pp. 469–483, Feb. 2007.
- [25] X. Zheng, X. Sun, K. Fu, and H. Wang, "Automatic annotation of satellite images via multifeature joint sparse coding with spatial relation constraint," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 4, pp. 652–656, Jul. 2013.
- [26] X. Yuan, X. Liu, and S. Yan, "Visual classification with multi-task joint sparse representation," *IEEE Trans. Image Process.*, vol. 21, no. 10, pp. 4349–4360, Oct. 2012.
- [27] J. Waqas, Z. Yi, and L. Zhang, "Collaborative neighbor representation based classification using  $\ell_2$ -minimization approach," *Pattern Recog. Lett.*, vol. 34, no. 2, pp. 201–208, Jan. 2012.
- [28] D. Baron, M. F. Duarte, M. B. Wakin, S. Sarvotham, and R. G. Baraniuk, "Distributed compressive sensing," 2009, arXiv preprint arXiv:1204.2358.
- [29] 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.
- [30] J. Li, H. Zhang, Y. Huang, and L. Zhang, "Hyperspectral image classification by nonlocal joint collaborative representation with a locally adaptive dictionary," *IEEE Trans. Geosci. Remote Sens.*, 10.1109/TGRS.2013.2274875, to be published.
- [31] M. Yang, L. Zhang, D. Zhang, and S. Wang, "Relaxed collaborative representation for pattern classification," in *Proc. IEEE Conf. CVPR*, 2012, pp. 2224–2231.
- [32] I. Jolliffe, *Principal Component Analysis*. Hoboken, NJ, USA: Wiley, 2005.
- [33] E. Angelopoulou, S. W. Lee, and R. Bajcsy, "Spectral gradient: A material descriptor invariant to geometry and incident illumination," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, 1999, pp. 861–867.



- [34] A. C. Bovik, M. Clark, and W. S. Geisler, "Multichannel texture analysis using localized spatial filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 1, pp. 55–73, Jan. 1990.
- [35] M. Shi and G. Healey, "Hyperspectral texture recognition using a multiscale opponent representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 5, pp. 1090–1095, May 2003.
- [36] T. C. Bau, S. Sarkar, and G. Healey, "Hyperspectral region classification using a three-dimensional Gabor filterbank," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 9, pp. 3457–3464, Sep. 2010.
- [37] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 309–320, Feb. 2001.
- [38] J. A. Benediktsson, M. Pesaresi, and K. Amason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1940–1949, Sep. 2003.
- [39] 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.
- [40] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.
- [41] G. M. Foody, "Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 5, pp. 627–634, May 2004.
- [42] R. Tibshirani, "Regression shrinkage and selection via the lasso: A retrospective," *J. R. Stat. Soc. B*, vol. 73, no. 3, pp. 273–282, Jun. 2011.
- [43] A. Y. Yang, S. S. Sastry, A. Ganesh, and Y. Ma, "Fast  $\ell_1$ -minimization algorithms and an application in robust face recognition: A review," in *Proc. 17th IEEE ICIP*, 2010, pp. 1849–1852.
- [44] 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.
- [45] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, May 2011. [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [46] J. A. Gualtieri and R. F. Crompt, "Support vector machines for hyperspectral remote sensing classification," in *Proc. 27th AIPR Workshop Adv. Comput.-Assisted Recog.*, 1998, pp. 221–232.



**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 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 the Head of the Remote Sensing Division, State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University. He is also a "Chang-Jiang Scholar" Chair Professor appointed by the Ministry of Education, China. He is currently a 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. He also serves as an Associate Editor of the *International Journal of Ambient Computing and Intelligence*, *International Journal of Image and Graphics*, *International Journal of Digital Multimedia Broadcasting*, *Journal of Geo-spatial Information Science*, and *Journal of Remote Sensing*. He has more than 300 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.

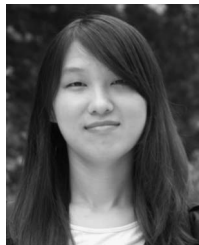
Dr. Zhang is a Fellow of IEEE, Executive Member (Board of Governor) of the China National Committee of International Geosphere–Biosphere Programme, and Executive Member of China Society of Image and Graphics. He regularly serves as a Cochair of a series of SPIE Conferences on Multispectral Image Processing and Pattern Recognition, Conference on Asia Remote Sensing, and many other conferences. He edits several conference proceedings, issues, and geoinformatics symposiums.



**Xin Huang** (M'13) received the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing (LIESMARS), Wuhan University, Wuhan, China, in 2009.

He is currently a Full Professor with LIESMARS. He has published more than 35 peer-reviewed articles in international journals. He has frequently served as a referee for many international journals for remote sensing. His research interests include hyperspectral data analysis, high-resolution image processing, pattern recognition, and remote sensing applications.

Dr. Huang was the recipient of the Top-Ten Academic Star of Wuhan University in 2009. In 2010, he received the Boeing Award for the best paper in image analysis and interpretation from the American Society for Photogrammetry and Remote Sensing. In 2011, he was the recipient of the New Century Excellent Talents in University from the Ministry of Education of China. In 2011, he was recognized by the IEEE Geoscience and Remote Sensing Society as the Best Reviewer of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. In 2012, he was the recipient of the National Excellent Doctoral Dissertation Award of China.



**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.

Dr. Zhang is a Reviewer of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, and so on.



**Lefei Zhang** (S'11–M'14) received the B.S. degree in sciences and techniques of remote sensing and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2008 and 2013, respectively.

In August 2013, he joined the Computer School, Wuhan University, where he is currently an Assistant Professor. His research interests include hyperspectral data analysis, high-resolution image processing, and pattern recognition in remote sensing images.

Dr. Zhang is a Reviewer of more than ten international journals, including IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, *Information Sciences*, and *Pattern Recognition*.