

Locality Regularized Robust PCRC: A Novel Simultaneous Feature Extraction and Classification Framework for Hyperspectral Images

Zhijing Yang, *Member, IEEE*, Faxian Cao, *Student Member, IEEE*, Yongqiang Cheng, Wing-Kuen Ling, *Senior Member, IEEE*, Ruo Hu

Abstract—Despite the successful applications of probabilistic collaborative representation classification (PCRC) in pattern classification, it still suffers from two challenges when being applied on hyperspectral images (HSIs) classification: (1) ineffective feature extraction in HSIs under noisy situation; (2) lack of prior information for HSIs classification. To tackle the first problem existed in PCRC, we impose the sparse representation to PCRC, i.e. to replace the 2-norm with 1-norm for effective feature extraction under noisy condition. In order to utilize the prior information in HSIs, we firstly introduce the Euclidean distance (ED) between the training samples and the testing samples for the PCRC to improve the performance of PCRC. Then we bring the coordinate information (CI) of the HSIs into the proposed model, which finally leads to the proposed locality regularized robust PCRC (LRR-PCRC). Experimental results show the proposed LRR-PCRC outperformed PCRC and other state-of-the-art pattern recognition and machine learning algorithms.

Index Terms—Coordinate information, Euclidean distance, hyperspectral image (HSIs), probabilistic collaborative representation classification (PCRC), sparse representation

I. INTRODUCTION

Hyperspectral images (HSIs) data has high spectral resolution due to the large spectral range and thus facilitate the discriminative performance on different land cover types [1]. The supervised classification [2] is one of the main topics in HSIs processing. In the supervised classification scenario, the output of testing samples is determined by the given training samples labelled for each class [1]. The great challenges, however, posed by adopting supervised

classification on HSIs are due to various causes. Firstly, the ratio of the large number of spectral bands in respect to the limited samples of training pixels is high, i.e. the well-known Hughes phenomenon [3]. Secondly, the materials of the same category may appear different spectral features and the different classes of samples may share the same spectral characteristic due to the sensory or environmental noises [4-5].

In order to address the Hughes phenomenon in hyperspectral images or remote sensing images, many advanced supervised classification methods have been proposed and achieved good performance, such as the support vector machine (SVM) [5], extreme learning machine (ELM) [6-7], sparse multinomial logistic regression (SMLR) [8-9] etc. In additions, some recent literatures have also been proposed for solving the Hughes Phenomenon. For examples, in [10] and [11], the band selection techniques have been proposed for reducing the spectral bands via adaptive subspace partition strategy and optimal clustering framework, respectively; in [12] and [13], a subpixel component analysis and recurrent attention technique have been proposed for hyperspectral image and scene classification, respectively; in [14] and [15], the performance of HSIs classification has been improved by multilabel sample augmentation-based and superpixel-based semi-supervised active learning, respectively. The above references have acquired relative good performances. On the other hand, methods such as sparse representation classification (SRC) [16-17] and collaborative representation classification (CRC) [1] also achieved good performance in HSIs classifications. Different from SVM, ELM or SMLR, the SRC and CRC don't need the training process because they aim to represent a testing sample directly by the linear combination of all the training samples from all classes with 1-norm and 2-norm on the representation coefficient, respectively [18-19]. Both methods finally assign the class labels to the corresponding testing samples directly via evaluating the minimum redundancy representation among all the classes [1]. This advantage of SRC and CRC has drawn vast research attentions. However, the SRC has its drawback in HSIs processing. The SRC is time-consuming as the computation of the sparsity 1-norm is a complicated procedure [1]. Meanwhile, Zhang et al. [19] pointed out that it is unnecessary to regularize the 1-norm coefficients if the feature dimension of samples is high enough [1]. They argued that the success of SRC can be largely attributed to the collaborative representation between the test sample and the training samples among all classes [19].

This work is supported in part by the National Nature Science Foundation of China (nos. 61471132, 61871136), the High-Level University Construction Funds of Guangdong University of Technology (no. 1109/220410011), the Innovation Team Project of Guangdong Education Department (no. 2017KCXTD011), the Major Project for Guangzhou Collaborative Innovation of Industry-University-Research (no.201704020196) and the China Scholarship Council (CSC). Corresponding author: Mr. Faxian Cao.

Z. Yang, F. Cao, W.-K Ling are with School of Information Engineering, Guangdong University of Technology, Guangzhou, 510006, China. (yzhj@gdut.edu.cn, faxiancao@foxmail.com, yongquanling@gdut.edu.cn).

F. Cao is also with the School of Electronic Engineering and Computer Science, Queen Mary, University of London, London, E1 4NS, UK. (faxian.cao@qmul.ac.uk).

Y. Cheng is with Department of Computer Science and Technology, University of Hull, Hull, HU6 7RX, UK. (y.cheng@hull.ac.uk).

R. Hu is with School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou, 510665, China. (hu68@163.com)

Therefore, CRC has attracted many attentions due to its good performance. For example, in [20], a relaxed CRC method has been proposed for multiple features fusion and classification with good performance. In [21], the good classification performance has been produced based on a collaborative neighbor representation method. In [22], a probabilistic collaborative representation classifier (PCRC) for pattern classification based on the viewpoint of probabilities and achieved better performance than CRC and some other methods [22]. Even for HSIs classification, the CRC also produced good performance. For example, a joint within-class collaborative representation has been proposed for hyperspectral image classification in [23]. In [24], a nonlocal joint collaborative representation with locally adaptive dictionary has been proposed for HSIs classification. In [25], a joint collaborative representation with multitask learning has been proposed for hyperspectral image classification. **These CRC-based methods have achieved good performance to some extent, but they ignored a critical issue of CRC that the 2-norm is sensitive to noise when HSIs are prone to be corrupted by noise as stated in literature [16].**

Noises may be introduced to HSIs during the process of HSIs data acquisition and transmission [26]. Hence, it is important to design an algorithm to extract features from noisy HSIs efficiently and effectively. For this purpose, many algorithms have been proposed, such as, graph-regularized low rank destriping (GRLD) [27], principal component analysis and wavelet shrinkage (PCA-WS) [28], spectral-spatial adaptive hyperspectral total variation (SSAHTV) [29] and singular spectrum analysis (SSA) [26] etc. **However, these methods focused on either feature extraction or denoising independently, resulting in the classification results that may not be optimal due to the lack of comprehensive consideration for both preprocessing and classification tasks.**

To address the above challenges, we propose a locality regularized robust probabilistic collaborative representation classification (LRR-PCRC) framework to process the feature extraction and classification simultaneously for HSIs classifications. Firstly, we introduce the sparse representation (SR) to the PCRC. That is because it is well-known that the 1-norm SR [30-31] is robust to characterize the loss function if the datasets are corroded by noises. Then, we introduce the rich prior information of HSIs to PCRC in order to extract the efficient feature from HSIs, including the Euclidean distance (ED) between training samples and testing samples, and the coordinate information (CI) of the HSIs. The motivations can be summarized in two aspects. On one hand, the same category shares similar spectral features to some extent even when the data points are corrupted, hence, the features can be extracted. More importantly, on the other hand, the neighborhood pixels within the regions are more likely to be in the same category [32]. Therefore, the coordinate information within HSIs can be used for extracting the spatial information since the coordinate of the pixels in HSIs can't be changed even the data points in HSIs are corrupted. **For the ED and CI information, we calculate the ED and coordinate distance between each training sample and the whole testing samples and add prior information**

we have acquired to impose the constraints to the representation coefficient. Finally, the locality regularization will be employed for robust HSIs feature extraction and classification simultaneously.

The main contributions of proposed work can be summarized as follows. Firstly, we introduce the sparse constraint to PCRC to tackle the problem that PCRC is sensitive to the noise. Secondly, the prior information in HSIs, Euclidean distance information and coordinate information between training samples and testing samples have been introduced to the PCRC for improving the classification results, leading to the proposed LRR-PCRC. Experimental results show the good performance of proposed LRR-PCRC in comparison with other state-of-the-art algorithms.

The remains of this work are structured as follows. Section II introduces the related works. The details of the proposed LRR-PCRC framework are discussed in Section III. Section IV shows the experimental results and the comprehensive analysis. Section V concludes this paper with insights of future work.

II. PROBABILISTIC COLLABORATIVE REPRESENTATION CLASSIFICATION (PCRC)

Given N training samples from K classes of a hyperspectral image: $X = [X_{N_1}, X_{N_2}, \dots, X_{N_K}] \in R^{d \times N}$, where d is the numbers of band in a hyperspectral image and X_{N_k} is the data matrix of the k -th class training samples ($N_1 + N_2 + \dots + N_K = N$). Let la_x and S denote the label set of all the classes in X and the linear subspace collaboratively spanned by the total samples in X , respectively [22]. Then given a data point $x \in R^{d \times 1}$ in the collaborative subspace S , it can be represented as follows: $x = X\alpha$, where $\alpha = [\alpha_{N_1}, \alpha_{N_2}, \dots, \alpha_{N_K}] \in R^{N \times 1}$ is the representation vector.

The PCRC [22] formulated S as a probabilistic collaborative subspace, as it argued that there should be different probabilities of $la_x \in la_x$ for a data point x . Once the 2-norm of α_{N_k} is small, the probability of $P(la_x \in la_{X_{N_k}})$ should be high [22]. It is intuitive to use a Gaussian function to define the probability, that is:

$$P(la_x \in la_{X_{N_k}}) \propto \exp(-b \| \alpha_{N_k} \|_2^2) \quad (1)$$

where b is a positive constant. Then the probability of every testing sample y that has the sample label of x can be formulated as $P(la_y = la_x)$. Hence, the probabilities that the testing sample y lies in the subspace S can be expressed as:

$$P(la_y \in la_x) = P(la_y = la_x/la_x \in la_x)P(la_x \in la_x) \quad (2)$$

where $P(la_y = la_x/la_x \in la_x) \propto \exp(-c \| y - x \|_2^2)$, c is a positive constant. Then Eq. (2) can be rewritten as:

$$\begin{aligned} P(la_y \in la_x) &\propto \exp(-(c \| y - X\alpha \|_2^2 + d \| \alpha \|_2^2)) \\ &= \exp(-(\| y - X\alpha \|_2^2 + \lambda \| \alpha \|_2^2)) \end{aligned} \quad (3)$$

where $\lambda = d/c$ is a constant. Since $x = X\alpha = \sum_{k=1}^K X_{N_k} \alpha_{N_k} = \sum_{k=1}^K x_k$, where x_k is equal to $X_{N_k} \alpha_{N_k}$ and can be seen as a

sample that belongs to the k -th class, then the probability that the testing sample y belongs to k -th class can be formulated as:

$$\begin{aligned} P(la_y = k) &= P(la_y = la_x/la_x = k)P(la_x = k) = \\ &P(la_y = la_x/la_x = k)P(la_x = k/la_x \in la_x)P(la_x \in la_x) \end{aligned} \quad (4)$$

where $P(la_x = k/la_x \in la_x)$ can be formulated as:

$$P(la_x = k/la_x \in la_x) \propto \exp(-\beta \|x - x_k\|_2^2) \quad (5)$$

where β is a positive constant.

Since $P(la_y = la_x/la_x \in la_x)$ is independent of the k -th class, once k belongs to la_x , i.e. $P(la_y = la_x/la_x \in la_x) = P(la_y = la_x/la_x = k) \propto \exp(-c \|y - x\|_2^2)$, then $P(la_y = k)$ can be expressed as:

$$\begin{aligned} &P(la_y = k) \\ &= P(la_y \in la_x)P(la_x = k/la_x \in la_x)P(la_y = la_x/la_x = k) \\ &\propto \exp(-(\|y - x\|_2^2) + \lambda \| \alpha \|_2^2 + \beta \|x - x_k\|_2^2) = \\ &\exp(-(\|y - X\alpha\|_2^2) + \lambda \| \alpha \|_2^2 + \beta \|X\alpha - X_{N_k}\alpha_{N_k}\|_2^2) \end{aligned} \quad (6)$$

Assuming the probabilities of $P(la_y = k)$ are independent, then all kinds of probabilities ($k = 1, 2, \dots, K$) can be computed by:

$$\begin{aligned} &\max P(la_y = 1, la_y = 2, \dots, la_y = K) \\ &= \max \prod_k P(la_y = k) \propto \exp(-(\|y - X\alpha\|_2^2 + \lambda \| \alpha \|_2^2 \\ &\quad + \beta / K \sum_{k=1}^K \|X\alpha - X_{N_k}\alpha_{N_k}\|_2^2)) \end{aligned} \quad (7)$$

Then ignoring the constant term after applying the logarithmic operator to the above Eq., we can have:

$$\begin{aligned} \hat{\alpha} &= \arg \min_{\alpha} \{ \|y - X\alpha\|_2^2 + \lambda \| \alpha \|_2^2 \\ &\quad + \beta / K \sum_{k=1}^K \|X\alpha - X_{N_k}\alpha_{N_k}\|_2^2 \} \end{aligned} \quad (8)$$

The solution of Eq. (8) can be expressed as:

$$\hat{\alpha} = (X^T X + \lambda I + \beta / K \sum_{k=1}^K \bar{X}_{N_k}^T \bar{X}_{N_k})^{-1} X^T y \quad (9)$$

where $\bar{X}_{N_k} = X - X_{N_k}$, $\bar{X}_{N_k} = [0, 0, \dots, X_{N_k}, \dots, 0] \in R^{d \times N}$.

Finally, the label of testing sample y can be formulated as:

$$la_y = \arg \max_k \exp(-\|X\hat{\alpha} - X_{N_k}\hat{\alpha}_{N_k}\|_2^2) \quad (10)$$

III. THE PROPOSED FRAMEWORK

A. LRR-PCRC Model

Although the PCRC has achieved good performance, there still exist some problems. As mentioned above, PCRC is sensitive to noises. The performance of PCRC will degrade if HSIs contain much noise. In order to address this drawback, we propose the LRR-PCRC for HSIs feature extraction and classification simultaneously. For N training samples from K classes of a hyperspectral image $X = [X_{N_1}, X_{N_2}, \dots, X_{N_K}] \in R^{d \times N}$ and n testing samples $Y = [Y_1^*, Y_2^*, \dots, Y_n^*] \in R^{d \times n}$, we need to represent the n testing samples Y using N training samples and assume both of them are corrupted by noise, that is

$$\begin{aligned} Y + E_1 &= (X + E_2)A = XA + E_2A \\ \xrightarrow{\Delta} Y &= XA + (E_2A - E_1) \xrightarrow{\Delta} Y = XA + E \end{aligned} \quad (11)$$

where $E_1 \in R^{d \times n}$, $E_2 \in R^{d \times N}$ and $E \in R^{d \times n}$ are noise, $A \in R^{N \times n}$ is the representation coefficient. From Eq. (11), we can see that the coefficient A needs to represent the n testing samples Y using N training samples under the situation that both of them are corrupted. In order to address this problem, we can formulate the PCRC to the following structure:

$$\begin{aligned} \hat{A} &= \arg \min_A \{ 1/2 \|Y - XA\|_F^2 + \lambda \|A\|_1 \\ &\quad + \beta / K \sum_{k=1}^K \|XA - X_{N_k}A_{N_k}\|_F^2 \} \end{aligned} \quad (12)$$

In Eq. (12), we replace the 2-norm with 1-norm considering that the 1-norm is robust to noise [31]. Furthermore, the feature extraction performance can be improved if we incorporate the rich prior information in HSIs to the model, thus producing better classification results. Recall the ED between training samples and testing samples, the spectral signal will be similar if the training samples and testing samples belong to the same class. Hence, we incorporate this information in Eq. (12), and then the model can be rewritten as:

$$\begin{aligned} \hat{A} &= \arg \min_A \{ 1/2 \|Y - XA\|_F^2 + \lambda \|\Gamma \odot A\|_1 \\ &\quad + \beta / K \sum_{k=1}^K \|XA - X_{N_k}A_{N_k}\|_F^2 \} \end{aligned} \quad (13)$$

where \odot is Hadamard operator [25], $\Gamma \in R^{N \times n}$ and $\Gamma_{i,j} = \|Y_j^* - X_i\|_2^2$ ($i = 1, 2, \dots, N; j = 1, 2, \dots, n$). Recall the coefficient A , each column of A , $A_{:,j} \in R^{N \times 1}$, represents the similarity between the testing pixel Y_j and the whole training samples X . The probability of the testing pixel Y_j belonging to the k -th class should be higher when the 2-norm of coefficient $A_{N_k,j}$ ($\sqrt{\sum_{N_i=1}^{N_k} |A_{N_i,j}|^2}$) is smaller which has been explained in Eq. (1). Also, we can deduce that the probability of the testing pixel Y_j belonging to the k -th class should be higher when the

1-norm of coefficient $A_{N_k,j}$ ($\sqrt{\sum_{N_i=1}^{N_k} |A_{N_i,j}|}$) is smaller since the 1-norm of coefficient $A_{N_k,j}$ is similar to the 2-norm of coefficient $A_{N_k,j}$. On the other hand, the 1-norm sparse representation is robust to noise as reported in [31].

Furthermore, we use the ED between the testing samples and training samples Γ to enhance the coefficient A , which can improve the classification performance both in terms of accuracies and robustness. Other than the ED between the testing samples and the training samples, the CI of training samples and testing samples can also enhance the coefficient A . **Firstly, we calculate the coordinate distances between each training sample and the whole testing samples, then the coordinate information which contains the locality information of training samples and testing samples can be used for enhancing the performance of HSIs classification.** This can be explained from two aspects: first, the probabilities that two samples belonging to the same class will be larger if they are

located closely [4]; second, even the data points in HSIs are corrupted by noise, the coordinate information of each pixel in HSIs remain the same. Hence, Eq. (13) can be represented as:

$$\hat{A} = \arg \min_A \left\{ \frac{1}{2} \|Y - XA\|_F^2 + \frac{\beta}{K} \sum_{k=1}^K \|XA - X_{N_k} A_{N_k}\|_F^2 + \lambda \|\Gamma \odot A\|_1 + \gamma \|C \odot A\|_1 \right\} \quad (14)$$

where $C \in R^{N \times n}$ and $C_{i,j} = (CC_{i,j}CC_{i,j}) / (\sum_{i=1}^N CC_{i,j}CC_{i,j})$, $CC_{i,j}$ can be expressed as follows:

$$CC_{i,j} = (|h_i - h_j| + |w_i - w_j|)^f \quad (15)$$

where h_i and w_i are the abscissa and ordinate of sample in the image, respectively. f is the smooth parameter that adjusts the distance decay speed. Fig. 1 shows the flowchart of the proposed LRR-PCRC.

B. Solutions for LRR-PCRC

In this subsection, we will derive the optimization algorithm to solve the proposed LRR-PCRC model based on the inexact augmented Lagrange Multiplier (IALM) [33]. Recall that the proposed LRR-PCR has three contributions compared with the PCRC, i.e. SR, ED between the training samples and testing sample, CI of the HSIs. Now we derive the solution for the proposed LRR-PCRC as follows (The other solutions of the PCRC with SR or ED, PCRC with SR and ED will be shown in Supplementary).

First, we introduce two auxiliary variables H and J to convert Eq. (14) by splitting the variable:

$$\min_{A,H,J} \left\{ \frac{1}{2} \|Y - XA\|_F^2 + \frac{\beta}{K} \sum_{k=1}^K \|XA - X_{N_k} A_{N_k}\|_F^2 + \lambda \|\Gamma \odot H\|_1 + \gamma \|C \odot J\|_1 \right\} \quad (16)$$

s. t. $H = A, J = A$.

Then, the corresponding augmented Lagrangian function for Eq. (16) can be rewritten as:

$$\min_{A,H,J} \left\{ \frac{1}{2} \|Y - XA\|_F^2 + \frac{\beta}{K} \sum_{k=1}^K \|XA - X_{N_k} A_{N_k}\|_F^2 + \lambda \|\Gamma \odot H\|_1 + \gamma \|C \odot J\|_1 + \langle Y_1, H - A \rangle + \langle Y_2, J - A \rangle \right\}$$

$$+ \frac{\tau}{2} (\|H - A\|_F^2 + \|J - A\|_F^2) \quad (17)$$

where Y_1 and Y_2 are the Lagrange multipliers. Then the alternative optimization algorithm [34] can be applied to solve the model of Eq. (17). Details are introduced as follows.

Update H: Fix A and J , then the H can be updated as:

$$H^{t+1} = \min_H \left\{ \frac{\lambda}{\tau^t} \|\Gamma \odot H^t\|_1 + \frac{1}{2} \|H^t - A^t + Y_1^t / \tau^t\|_F^2 \right\} \quad (18)$$

The solution of Eq. (18) can be solved by the transformation of the soft-threshold rule [35]:

$$\begin{aligned} H^{t+1} &= \text{soft} \left(A^t - \frac{Y_1^t}{\tau^t}, \Gamma \times \lambda / \tau^t \right) \\ &= \max \{0, \text{abs}(e) - (\Gamma \times \lambda / \tau^t)\} \times \text{sign}(e) \end{aligned} \quad (19)$$

where $e = A^t - Y_1^t / \tau^t$.

Update J: Fix A and H , J can be updated as:

$$J^{t+1} = \min_J \left\{ \frac{\gamma}{\tau^t} \|C \odot J\|_1 + \frac{1}{2} \|J^t - A^t + Y_2^t / \tau^t\|_F^2 \right\} \quad (20)$$

The solution of Eq. (20) can be computed as:

$$\begin{aligned} J^{t+1} &= \text{soft} \left(A^t - \frac{Y_2^t}{\tau^t}, \Gamma \times \lambda / \tau^t \right) \\ &= \max \{0, \text{abs}(e) - (\Gamma \times \lambda / \tau^t)\} \times \text{sign}(e) \end{aligned} \quad (21)$$

where $e = A^t - Y_2^t / \tau^t$.

Update A: Fix H and J , A and be updated as:

$$\begin{aligned} A^{t+1} &= \min_A \left\{ \frac{1}{2} \|Y - XA^t\|_F^2 + \frac{\beta}{K} \sum_{k=1}^K \|XA^t - X_{N_k} A_{N_k}^t\|_F^2 + \frac{\tau}{2} (\|H^t - A^t + Y_1^t / \tau^t\|_F^2 + \|J^t - A^t + \frac{Y_2^t}{\tau^t}\|_F^2) \right\} \end{aligned} \quad (22)$$

Then the solution of Eq. (22) can be achieved by the first-order derivation:

$$A^{t+1} = \left(X^T X + \frac{\beta}{K} \sum_{k=1}^K \bar{X}_{N_k}^T \bar{X}_{N_k} \right)^{-1} \left(X^T Y + \tau^t (H^t + J^t) + Y_1^t + Y_2^t \right) \quad (23)$$

Finally, the overall optimization problem for solving the proposed LRR-PCRC is summarized in Algorithm 1.

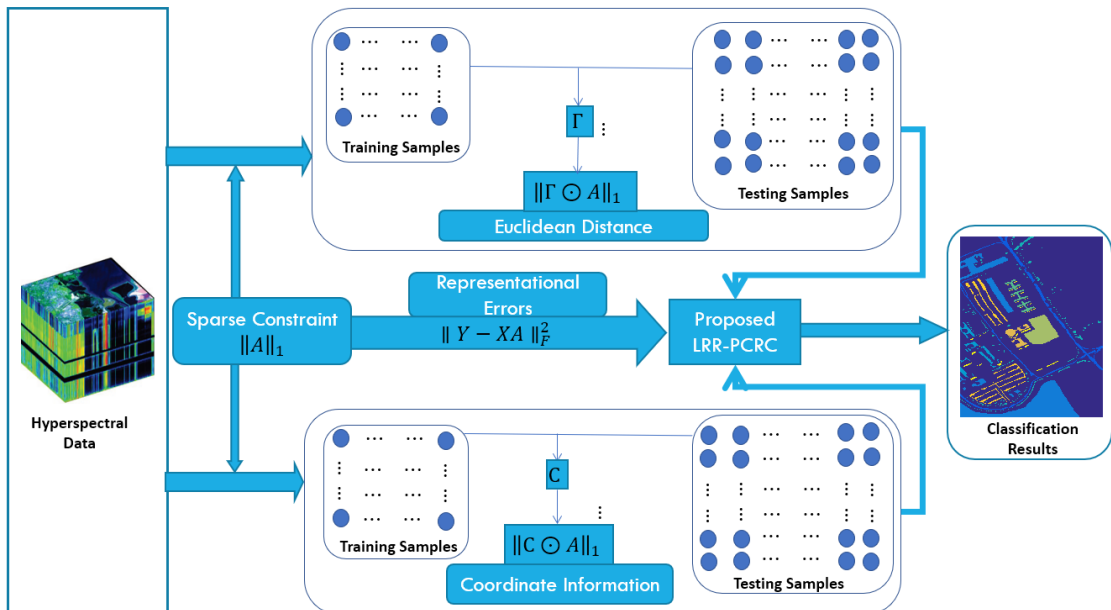


Fig. 1 The classification flowchart of proposed LRR-PCRC algorithm.

Algorithm 1: The proposed LRR-PCRC

Input: For N training samples from K classes of a hyperspectral image $X = [X_{N_1}, X_{N_2}, \dots, X_{N_K}] \in R^{d \times N}$ and n testing samples $Y = [Y_1, Y_2, \dots, Y_n] \in R^{d \times n}$, the parameters $\beta, \lambda, \gamma, f, \tau^0 = 10 * \lambda, Y_1 = 0, Y_2 = 0$.

1.1 Set $t=0$

1.2 Update H :
 $H^{t+1} = \arg \min_H \lambda / \tau^t \|H^t\|_1 + 1/2 \|H^t - A^t + Y_1^t / \tau^t\|_F^2 \Rightarrow H^{t+1} \leftarrow \max\{0, \text{abs}(A^t - \frac{Y_1^t}{\tau^t}) - (\Gamma \times \lambda / \tau^t)\} \times \text{sign}(A^t - Y_1^k / \tau^t)$

1.3 Update J :
 $J^{t+1} = \min_J \gamma / \tau^t \|C \odot J\|_1 + 1/2 \|J^t - A^t + Y_2^t / \tau^t\|_F^2 \Rightarrow J^{t+1} = \max\{0, \text{abs}(A^t - \frac{Y_2^t}{\tau^t}) - (\Gamma \times \gamma / \tau^t)\} \times \text{sign}(A^t - Y_2^k / \tau^t)$

1.4: Update A :
 $A^{t+1} = \arg \min_A 1/2 \|Y - XA^t\|_F^2 + \beta / K \sum_{k=1}^K \|XA^t - X_{N_k} A_{N_k}^t\|_F^2 + \tau^t / 2 (\|H^t - A^t + Y_1^t / \tau^t\|_F^2 + \|J^t - A^t + \frac{Y_2^t}{\tau^t}\|_F^2)$
 $\Rightarrow A^{t+1} = (X^T X + \beta / K \sum_{k=1}^K \bar{X}_{N_k}^T \bar{X}_{N_k})^{-1} (X^T Y + \tau^t (H^t + J^t) + Y_1^t + Y_2^t)$

1.5 Update other parameters:
 $Y_1^{t+1} = Y_1^t + \tau^t (H^{t+1} - A^{t+1}); Y_2^{t+1} = Y_2^t + \tau^t (J^{t+1} - A^{t+1}), \tau^{t+1} = 1.1 \times \tau^t$.

1.6 Quit the algorithm if the stopping criterion is met; otherwise, go back to Step 1.2.

Output: Predict the testing sample label:
 $la_{Y_i^*} = \arg \max_k \exp(-\|X\hat{\alpha} - X_{N_k} \hat{\alpha}_{N_k}\|_2^2)$ or $la_{Y_i^*} = \arg \min_k (\|X\hat{\alpha} - X_{N_k} \hat{\alpha}_{N_k}\|_2^2); i = 1, \dots, n, k = 1, \dots, K$

TABLE I. THE ABBREVIATIONS OF THIS PAPER

Abbreviation List			
PCRC	probabilistic collaborative representation classification	SSA	singular spectrum analysis
HSIs	hyperspectral images	LRR-PCRC	locality regularized robust-probabilistic collaborative representation classification
ED	Euclidean distance	IALM	inexact augmented Lagrange multiplier
CI	coordinate information	SVM-CK	SVM with composite kernel
LRR	locality regularized robust	AP	attribute profile
SVM	support vector machine	KPI	key performance indicators
ELM	extreme learning machine	OA	overall accuracies
SMLR	sparse multinomial logistic regression	AA	average accuracies
SRC	sparse representation classification	k	kappa coefficient
CRC	collaborative representation classification	CA	category accuracies
GRLD	graph-regularized low rank destriping	Tr	training time
PCA-WS	principal component analysis and wavelet shrinkage	Ts	testing time
Val	Validation	RF	Random Forest
RoF	Rotation Forest	RoRF-KPCA	rotation random forest-kernel principal component analysis
SSAHTV	spectral-spatial adaptive hyperspectral total variation		

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Datasets Description

(1) Indian Pines dataset

The Indian Pines [26] dataset has been collected by AVIRIS sensors in June 1992 which consists of urban images. The 20 heavily noise bands have been removed since these bands are severely affected by the water absorptions and thus 200 bands remained. The spectral range varies from 0.2µm to 2.4µm while the spatial resolution is 20m per pixel with size of 145×145 pixels. There are in total 10366 samples within 16 classes need to be classified in this dataset.

(2) Pavia University dataset

The Pavia University [26] dataset has been collected by ROSIS sensors over the area surrounding the University of Pavia, Italy in 2001. The 12 heavily noise bands have been removed since these bands are severely affected by the water absorptions and thus 103 bands remained. There are in total 42776 labelled samples within 9 classes available for classification in this dataset.

B. Benchmarking Approaches

To validate our proposed LRR-PCRC, we have compared LRR-PCRC with the state-of-the-art methods to benchmark the performance of our method, including the SMLR [9], SVM [36] and SVM with composite kernel (SVM-CK), SMLR with attribute profile (SMLR-AP) [9], PCRC and PCRC-AP. The

Matlab codes of SVM and SVM-CK are downloadable from [37], while the code of SMLR and AP are downloadable from [38]. The parameters of AP are set according to recommendation in [9].

The experiments are carried out on a computer with 2.9GHz i7 7820HQ CPU with 32GB RAM running win10 OS. The codes are written in Matlab (R2015a) **and all the experimental results in this paper are 10 times average.**

The key performance indicators (KPI) include overall accuracies (OA), average accuracies (AA), kappa coefficient (κ), category accuracies (CA), training time (Tr) and testing time (Ts). **In addition, all the abbreviations in this paper have been displayed in Table 1.**

C. Parameter Analysis

The key parameters for PCRC are the β and λ in Eq. (9) and additional parameters for proposed LRR-PCRC are f in Eq. (15), γ in Eq. (14). Four experiments are carried out to evaluate the parameters of λ , β , γ and f , respectively, using 5 training samples per class. Since the proposed LRR-PCRC focuses on the feature extraction and classification, we consider two different situations of the raw data, i.e. without additional noise and with normally distribution noise (*i.i.d.*: zero mean with σ^2 covariance). In this subsection, we set σ to be 0.02. We fix the other parameters when evaluating one parameter. The experimental parameters setting for this subsection can be seen at Table 2. In order to achieve a reliable result, we have iterated each experiment 10 times and averaged the results.

Experiment 1 (λ): In this experiment, we evaluate the parameter $\lambda = 2^{a1}$ which $a1$ ranges from [-20, -19, ..., 5]. $\beta = 2^{-10}$ for both PCRC and LRR-PCRC. $f = 3$ and $\gamma = 2^{12}$ for LRR-PCRC. It can be seen from Fig. 2(a) and (b) that λ has some impact in both PCRC and LRR-PCRC. In addition, in Indian Pines and Pavia University datasets, the OA of the proposed LRR-PCRC is higher than PCRC under the situation with or without addition noise. We also can see that the OA of PCRC decreased dramatically when add the noise to HSIs data. However, the proposed LRR-PCRC is more robust to noise than PCRC.

As shown in the Fig. 2(a) and (b), the PCRC achieved best accuracies at about $\lambda = 2^{-7}$ and $\lambda = 2^{-2}$ under the situation with and without noise respectively in Indian Pines. And in Pavia University dataset, the PCRC achieve good classification accuracies at $\lambda = 2^{-7}$ for both no additional noise and noise condition. Hence, above parameters have been selected for PCRC if not otherwise mentioned. For the proposed LRR-PCRC, the classification accuracies are relatively stable ranging from $\lambda = 2^{-15}$ to $\lambda = 2^{-5}$ in Indian Pines dataset while ranging from $\lambda = 2^{-15}$ to $\lambda = 2^3$ in Pavia University dataset. Hence, the $\lambda = 2^{-10}$ and $\lambda = 2^0$ have been selected for Indian Pines and Pavia University datasets, respectively.

Experiment 2 (β): In this experiment, we evaluated the parameter $\beta = 2^{a2}$ which $a2$ ranges from [-20, -19, ..., 5]. It can be seen from Fig. 2 (c) and (d) that the β also has a certain impact on PCRC in both Indian Pines and Pavia University datasets, and on LRR-PCRC in Indian Pines dataset. In

addition, the proposed LRR-PCRC not only produces higher classification accuracies than PCRC, but also demonstrates more robustness to noise than PCRC. As also can be seen from Fig. 2 (c) and (d) that about at the value of $\beta = 2^{-10}$, PCRC has produced the best classification accuracies in Indian Pines dataset. And in Pavia University dataset, PCRC produced the best classification accuracies at $\beta = 2^{-7}$ and $\beta = 2^{-4}$ under the condition with and without additional noise respectively. Hence, above parameters have been selected for PCRC if not otherwise mentioned. For the proposed LRR-PCRC, it performs stably when β varies from [-20, ..., -5] and [-20, ..., 5] in Indian Pines dataset and Pavia University dataset, respectively. Hence, if not specially mentioned in the following experiments, β is set to 2^{-8} and 2^3 for the proposed LRR-PCRC in Indian Pines dataset and Pavia University dataset, respectively.

Experiment 3 (γ): In this experiment, we evaluated the parameter $\gamma = 2^{a3}$ which $a3$ ranges from [1, 2, ..., 15]. It can be seen from Fig. 2(e) and (f) that the OA of LRR-PCRC increased then decreased when γ increases. In addition, the noise has more impact in Indian Pines dataset than in Pavia University dataset when γ is small. However, the impact can be mostly eliminated when γ is set to a higher value. It would be good for classification performance when λ and β are set to small values while γ is set to a big value. This is because that the coordinate information may dominate the important role for feature extraction and classification. We can see that LRR-PCRC yield best classification accuracies both in Indian Pines dataset and Pavia University dataset at $\gamma = 2^{12}$ and $\gamma = 2^9$, respectively. Hence, $\gamma = 2^{12}$ and $\gamma = 2^9$ have been selected for LRR-PCRC.

Experiment 4 (f): In this experiment, we evaluated the parameter f which f ranges from [1, 2, ..., 10]. It can be seen that from Fig.2 (g) and (h) that the OA of LRR-PCRC increases then decreases when f increased. It can be also seen that the best value of f for classification accuracies is 3 in both dataset and both situations that with and without additional noise. Therefore $f = 3$ has been selected in the following experiments.

TABLE 2. THE PARAMETERS SETTING OF PCRC AND LRR-PCRC FOR PARAMETER ANALYSIS

Dataset	Methods	Experiment				
		1	2	3	4	
Indian Pines	No	PCRC	$\beta = 2^{-10}$	$\lambda = 2^{-7}$	-	-
	additional noise	LRR-PCRC	$\beta = 2^{-10}$	$\lambda = 2^{-10}$	$\lambda = 2^{-10}$	$\lambda = 2^{-10}$
			$f = 3$	$f = 3$	$\beta = 2^{-8}$	$\beta = 2^{-8}$
			$\gamma = 2^{12}$	$\gamma = 2^{12}$	$f = 3$	$\gamma = 2^{12}$
	Additional noise	PCRC	$\beta = 2^{-10}$	$\lambda = 2^{-2}$	-	-
		LRR-PCRC	$\beta = 2^{-10}$	$\lambda = 2^{-10}$	$\lambda = 2^{-10}$	$\lambda = 2^{-10}$
		$f = 3$	$f = 3$	$\beta = 2^{-8}$	$\beta = 2^{-8}$	
Pavia University	No	PCRC	$\beta = 2^{-10}$	$\lambda = 2^{-7}$	-	-
	additional noise	LRR-PCRC	$\beta = 2^{-10}$	$\lambda = 2^0$	$\lambda = 2^0$	$\lambda = 2^0$
			$f = 3$	$f = 3$	$\beta = 2^3$	$\beta = 2^3$
			$\gamma = 2^{12}$	$\gamma = 2^{12}$	$f = 3$	$\gamma = 2^9$
	Additional noise	PCRC	$\beta = 2^{-10}$	$\lambda = 2^{-4}$	-	-
		LRR-PCRC	$\beta = 2^{-10}$	$\lambda = 2^0$	$\lambda = 2^0$	$\lambda = 2^0$
		$f = 3$	$f = 3$	$\beta = 2^3$	$\beta = 2^3$	
		$\gamma = 2^{12}$	$\gamma = 2^{12}$	$f = 3$	$\gamma = 2^9$	

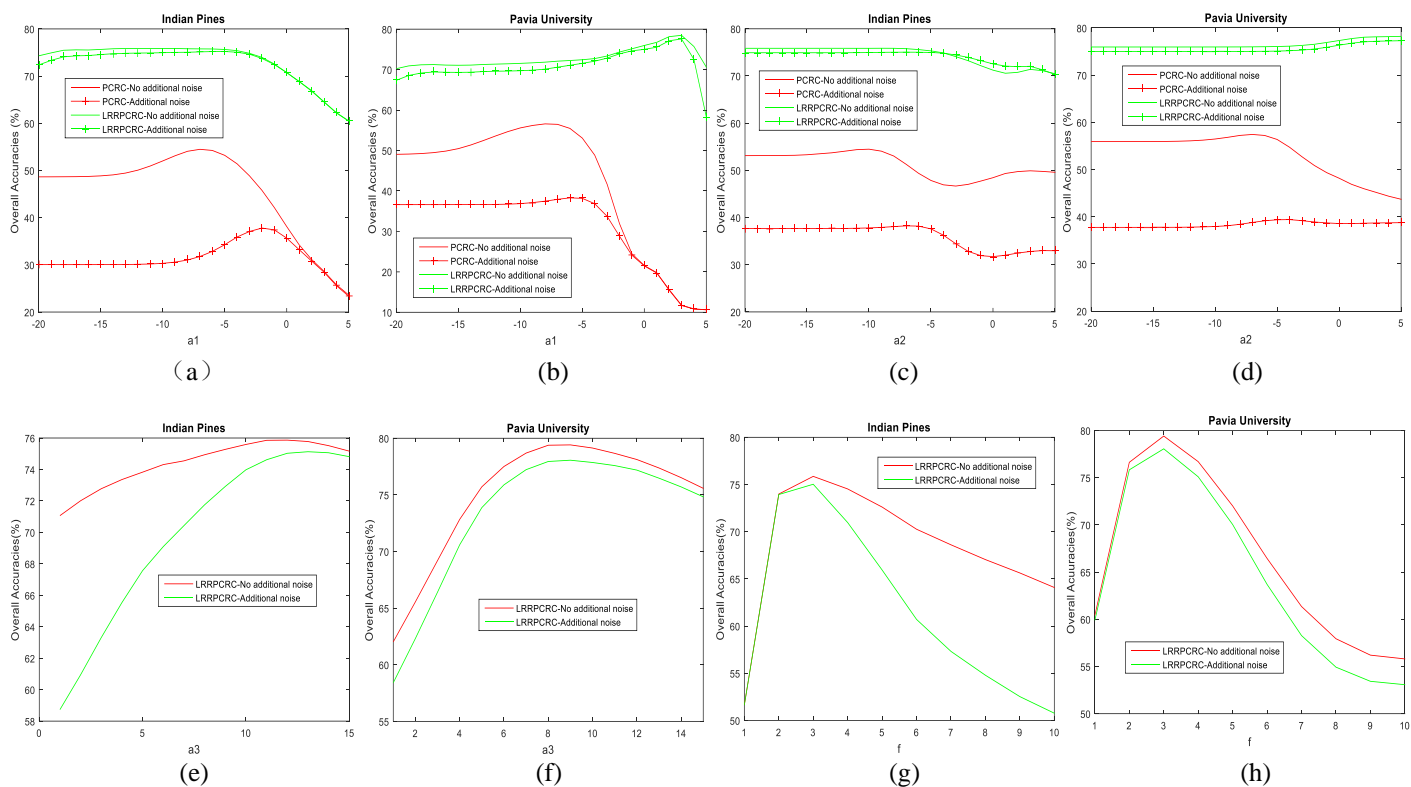


Fig. 2 The impact of PCRC and proposed LRR-PCRC without and with additional noise in Indian Pine and Pavia University. $\lambda/a1$ (a) and (b), $\beta/a2$ (c) and (d), $\gamma/a3$ (e) and (f), f (g) and (h).

D. Contribution Analysis

In this section, we analyze the three contributions of the proposed LRR-PCRC. Compared with PCRC, the proposed LRR-PCRC has three main contributions: 1) the LRR-PCRC with sparse representation for improving the feature extraction performance, 2) LRR-PCRC with SR and ED, and 3) LRR-PCRC with SR, ED and CI for extracting the efficient features of HSIs in order to further improve the classification accuracies. Hence, we will show the impact of each contribution of the proposed LRR-PCRC in this subsection. In order to show the exact impact of contributions on proposed model, we selected the best parameters of these contributions which can be seen in Table 3. We used 20 training samples per class under the condition with and without additional noise $\sigma = 0.02$ (i.i.d: zero mean with σ^2 covariance).

As can be seen from Table 4, three contributions have its improvement of classification accuracies in both Indian Pine dataset and Pavia University dataset. We can also see a phenomenon, ED information seems to be less important in Indian Pine dataset but important in Pavia University dataset. This can be explained by the reason that the Indian Pine and Pavia University dataset have different spatial structures. In more details, the pixels of the same category in Indian Pine dataset are located together, and the shapes of the same category are more likely to be blocked. While in Pavia University dataset, the shapes of some categories are more likely to be stripe, hence, the CI has a dominant role in Indian Pine dataset. However, in Pavia University dataset, both the

ED and CI are very important. From above discussion, it can be seen that it's more reasonable to use CI and ED collaboratively.

E. Effect of varying numbers of training samples and the varying level of noise

In this subsection, we will further evaluate the performance in feature extraction and classification of the proposed LRR-PCRC via varying the numbers of training sample and levels of noise. The training samples vary between 20, 25 and 30 while the noise levels vary between 0.2, 0.4, 0.6 and 0.8.

As can be seen from Tables 5 and 6, the classification results in Indian Pine dataset and Pavia University dataset by LRR-PCRC are more stable and better than PCRC when we vary the numbers of training sample and levels of noise. This further demonstrates the superior performance of the proposed LRR-PCRC.

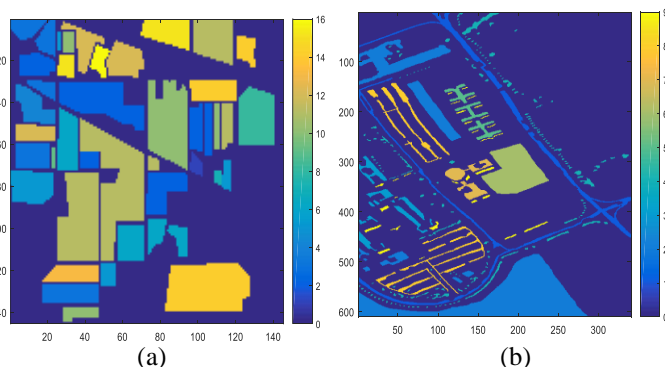


Fig. 3. (a) The ground truth for Indian Pine hyperspectral dataset; (b) The ground truth for Pavia University hyperspectral dataset.

TABLE 3. THE PARAMETERS SETTING OF THE CONTRIBUTIONS FOR CONTRIBUTIONS ANALYSIS

	Indian Pines			Pavia University		
	LRR-PCRC			LRR-PCRC		
	SR	SR+ED	SR+ED+CI	SR	SR+ED	SR+ED+CI
No noise	$\lambda = 2^{-6}$ $\beta = 2^{-15}$	$\lambda = 2^{-9}$ $\beta = 2^{-13}$	$\tau^0 = 2^{-10}; \beta = 2^{-8}$ $\gamma = 2^{12}; f = 3$	$\lambda = 2^{-10}$ $\beta = 2^{-15}$	$\lambda = 2^{-7}$ $\beta = 2^{-13}$	$\tau^0 = 2^{-10}; \beta = 2^{-6}$ $\gamma = 2^5; f = 3$
Noise	$\lambda = 2^{-4}$ $\beta = 2^{-9}$	$\lambda = 2^{-4}$ $\beta = 2^{-8}$	$\tau^0 = 2^{-10}; \beta = 2^{-6}$ $\gamma = 2^{12}; f = 3$	$\lambda = 2^{-7}$ $\beta = 2^{-20}$	$\lambda = 2^{-4}$ $\beta = 2^{-13}$	$\tau^0 = 2^{-10}; \beta = 2^{-4}$ $\gamma = 2^{13}; f = 3$

TABLE 4. CLASSIFICATION ACCURACY (%) AND STANDARD DEVIATION WITH 20 TRAINING SAMPLES PER CLASS FOR INDIAN PINES AND PAVIA UNIVERSITY DATASET (BEST RESULTS IN BOLD).

Index		Indian Pines					Pavia University				
		PCRC	LRR-PCRC				PCRC	LRR-PCRC			
			SR	SR+ED	SR+CI	SR+CI+ED		SR	SR+ED	SR+CI	SR+CI+ED
No noise	OA	61.2±2.8	68.0±2.7	71.5±2.4	91.0±0.7	91.0±0.8	62.0± 2.8	66.1±4.9	77.1± 2.3	86.8±3.2	90.4±1.2
	AA	72.4±1.9	70.8±2.5	83.6±0.9	93.9±0.7	94.1±0.5	68.8± 2.1	64.0± 2.7	83.0±0.8	88.7±1.7	93.5±0.6
	k	56.8±2.9	63.7±3.1	68.1±2.6	89.8± 0.8	89.7±0.9	51.4± 3.0	56.1±5.1	70.7±2.6	83.0±4.0	87.5± 1.5
Noise 0.02	OA	43.7±1.3	51.7±1.4	66.7±2.9	87.5±0.7	87.6±0.8	39.4± 3.4	44.1±6.0	65.1±2.7	84.5±2.1	88.7 ± 1.3
	AA	45.1±0.9	49.6± 1.5	77.6± 1.7	91.0±1.1	91.8±0.7	49.0± 1.0	44.4±1.7	74.1±0.9	83.6± 1.1	89.7 ± 0.7
	k	37.2±1.2	45.2±1.6	62.8±3.0	85.8±0.7	85.9±0.9	28.4 ±2.6	32.4± 5.1	56.4±2.9	80.0± 2.6	85.3 ± 1.6

TABLE 5. CLASSIFICATION ACCURACY (%) AND STANDARD DEVIATION IN INDIAN PINES

Index		No noise		Noise 0.02		Noise 0.04		Noise 0.06		Noise 0.08	
		PRRC	LRR-PCRC	PRRC	LRR-PCRC	PRRC	LRR-PCRC	PRRC	LRR-PCRC	PRRC	LRR-PCRC
20	OA	61.2±2.8	91.0±0.8	43.7±1.3	87.6±0.8	30.4±0.9	83.5±1.1	23.6±0.9	80.1± 1.4	19.8± 0.9	77.1± 1.7
	AA	72.4±1.9	94.1±0.5	45.1±0.9	91.8±0.7	32.6±1.3	87.3±1.1	25.8±1.3	82.6± 1.3	21.8±1.0	78.1± 0.9
	k	56.8±2.9	89.7±0.9	37.2±1.2	85.9±0.9	23.2±0.9	81.3±1.3	16.3± 0.8	77.5± 1.6	12.5± 0.7	74.3± 1.8
25	OA	63.4±2.5	91.9±1.1	45.4± 1.4	88.8±0.9	31.6± 1.1	84.7±1.0	25.1± 1.0	81.3±1.0	21.6±0.9	78.5±0.8
	AA	73.3±2.3	95.0±0.4	46.0±1.0	92.4±0.9	33.8± 0.8	87.2±1.3	27.0±0.7	81.9±1.7	23.1± 0.5	77.7±1.3
	k	59.2±2.5	90.8±1.2	39.0±1.2	87.2± 1.0	24.4± 0.9	82.6±1.1	17.7±0.7	78.8±1.1	14.2±0.6	75.7±0.9
30	OA	63.9±2.4	92.7±0.9	45.5± 0.9	89.59±0.6	31.8 ± 1.1	85.7±0.6	25.4±1.2	82.5±0.8	22.2±1.1	79.6±1.0
	AA	73.2±1.3	95.1± 0.8	45.8± 0.9	92.7±0.6	33.7 ± 0.9	87.4±1.7	27.4±1.1	82.3±1.4	24.0± 0.9	78.7±1.0
	k	59.7±2.4	91.7± 1.0	39.2±0.9	88.1±0.7	24.6 ± 1.0	83.7±0.7	18.0±1.0	80.1±0.9	14.8±0.9	76.9±1.1

TABLE 6. CLASSIFICATION ACCURACY (%) AND STANDARD DEVIATION WITH DIFFERENT TRAINING NUMBERS AND NOISE LEVELS IN PAVIA UNIVERSITY DATASET

Index		No noise		Noise 0.02		Noise 0.04		Noise 0.06		Noise 0.08	
		PRRC	LRR-PCRC	PRRC	LRR-PCRC	PRRC	LRR-PCRC	PRRC	LRR-PCRC	PRRC	LRR-PCRC
20	OA	62.0±2.8	90.4±1.2	39.4± 3.4	88.7 ± 1.3	33.5± 2.7	84.1±1.4	31.4± 2.5	79.4±1.7	30.2± 2.5	73.8±2.1
	AA	68.8±2.1	93.5±0.6	49.0± 1.0	89.7 ± 0.7	42.0± 1.1	80.9±0.4	38.9± 1.1	78.1±0.8	37.2± 1.1	74.7±1.2
	k	51.4±3.0	87.5±1.5	28.4 ±2.6	85.3 ± 1.6	22.3± 2.0	79.5±1.7	20.1± 1.9	73.8±2.0	18.9± 1.8	67.2±2.3
25	OA	63.4±1.7	91.6±1.7	44.2± 2.7	90.0± 2.0	37.7± 2.3	85.2±1.9	34.6± 1.9	81.1±1.7	32.8±1.6	76.0± 2.1
	AA	69.9±1.8	94.4±0.5	51.0± 1.2	91.0± 0.8	44.2± 1.0	81.0± 0.8	40.7± 1.0	78.4±0.9	38.7± 0.8	75.0±1.1
	k	53.0±1.7	89.1± 2.1	32.5± 2.3	87.0±2.5	26.0± 1.7	80.9± 2.4	23.0± 1.4	75.7±2.1	21.2± 1.2	69.7±2.4
30	OA	64.0±1.2	92.4±1.1	45.0± 3.0	90.4±1.5	39.5± 2.6	85.4±1.9	36.8± 2.4	80.8±2.3	34.9± 2.1	75.2±2.4
	AA	71.2±1.0	94.8±0.5	52.5± 1.0	91.3±0.7	45.4± 0.8	82.2±0.8	41.6± 0.8	79.6±1.4	39.4± 0.8	76.5±1.5
	k	53.8±1.4	90.1± 1.5	33.3± 2.6	87.5±1.9	27.4± 2.1	81.1±2.3	24.7± 2.0	75.5±2.8	22.8± 1.7	68.9± 2.7

F. Comparison with other state-of-the-art algorithms

In this subsection, firstly, we verified the proposed methods under 1% training samples, and the corresponding training samples and testing samples are reported in Table 7. The experiments are conducted in two conditions, i.e. no additional noise and noise with level of $\sigma = 0.2$. Tables 8 and 9 display the classification accuracies in Indian Pines and Pavia University dataset. As can be seen from Tables 8 and 9, the classification accuracies of the proposed LRR-PCRC are better than PCRC and other methods in the two conditions. This verifies the good performance of the proposed LRR-PCRC again. Figs. 4 and 5 show the corresponding classification map.

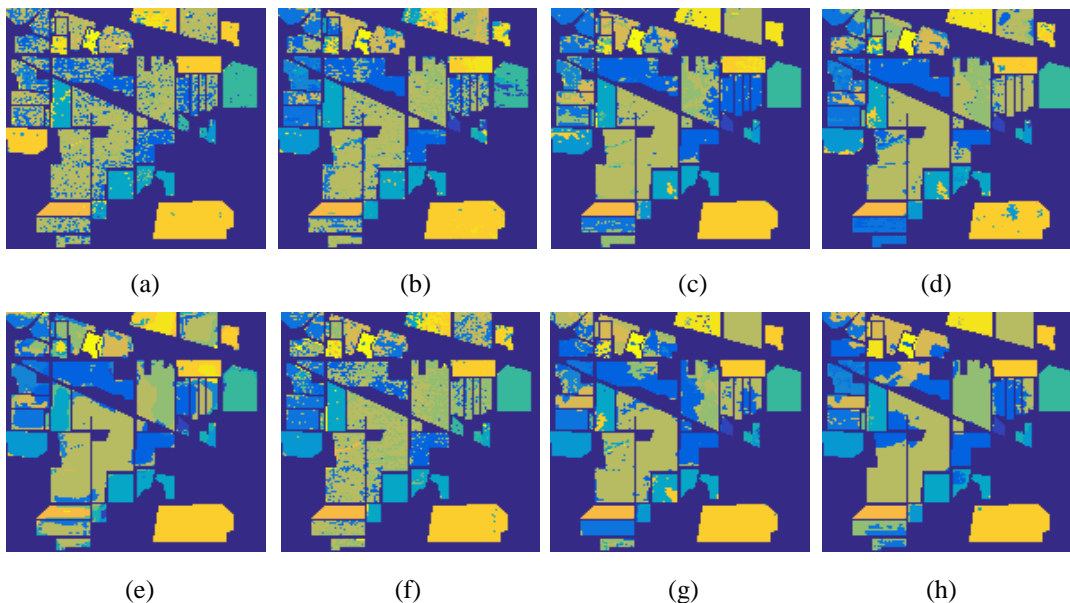
Secondly, to further verify the proposed methods, we have divided the whole hyperspectral dataset into three subsets (subset 1, 2 and 3) randomly with 10 times. In each time, the subset 1 and subset 2 are randomly selected 5 samples per class which will be used for two-fold cross validations and subset 3 containing the rest of the samples the is used for testing. We use subset 1 for training and the subset 2 for validation and then the subset 2 will be used for training and subset 1 will be utilized for validation. Finally, all the classification results will be averaged. Tables 10 and 11 show the results for training (Tr), validation (Val) and testing (Ts). As can be seen from these two tables, in general, the proposed method acquired better results than PCRC and other methods in both datasets.

Thirdly, Table 12 shows the time consumed in training and testing with 20 training samples per class without additional noise. It should be noted that the time costs of these methods would not be changed since the dimensionality of HSIs hasn't changed when noise are added to HSIs. From Table 12, we can see that in both Indian Pines and Pavia University dataset, the time consumed by our proposed LRR-PCRC is higher than SMLR, SVM. SMLR-AP and SVM-AP, but that of the proposed LRR-PCRC is less than PCRC and PCRC-AP. The proposed LRR-PCRC has higher time cost than SVM-CK in Pavia University dataset but lower in Indian Pines and Pavia University dataset. Now, let us theoretically analyze the computation complexity of the proposed LRR-PCRC to clarify

the reason that the proposed method has more computational time in comparison with other methods. As can be seen from Algorithm 1, the proposed LRR-PCRC has four steps (Step 1.2 to Step1.5) for updating. The major cost is the inverse operation in Step 1.4 and this is the main reason that the proposed method has high computational time compared with SVM-based and SMLR-based methods. The proposed LRR-PCRC needs to compute the inverse operation in each iteration. In Step 1.2, 1.3 and 1.5, the major cost is just matrix multiplication and addition, hence, the time consumed by these operations can be ignored comparing to the inverse operation. Given a square matrix with the size of $g \times g$, the computation complexity of inverse operation for such matrix is $O(g^3)$ and thus the computation complexity of our method is $O(N^3)$. Furthermore, we can see that the classification accuracies of the proposed LRR-PCRC are much higher than the other classifiers. Hence, we can conclude that the proposed LRR-PCRC has good performance in terms of above analysis.

TABLE 7. THE TRAINING/TESTING SAMPLES IN THE TWO DATASETS

Index/category	Indian Pines		Pavia University	
	Train	Test	Index/Category	Train Test
1 Alfalfa	3	51	1 Asphalt	66 6565
2 Corn-no till	14	1420	2 Meadows	186 18463
3 Corn-min till	8	826	3 Gravel	20 2079
4 Corn	4	230	4 Trees	30 3034
5 Grass/pasture	5	492	5 Metal sheets	13 1332
6 Grass/tree	8	739	6 Bare soil	50 4979
7 Grass/pasture-mowed	3	23	7 Bitumen	13 1317
8 Hay-windrowed	5	484	8 Bricks	37 3645
9 Oats	2	18	9 Shadows	10 937
10 Soybeans-no till	10	958		
11 Soybeans-min till	24	2444		
12 Soybeans-clean till	7	607		
13 Wheat	4	208		
14 Woods	13	1281		
15 Bldg-grass-tree-drives	5	375		
16 Stone-steel towers	4	91		



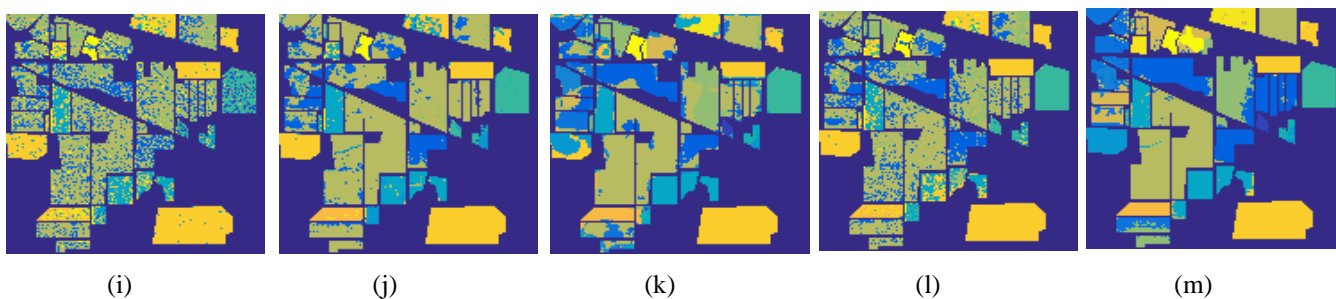
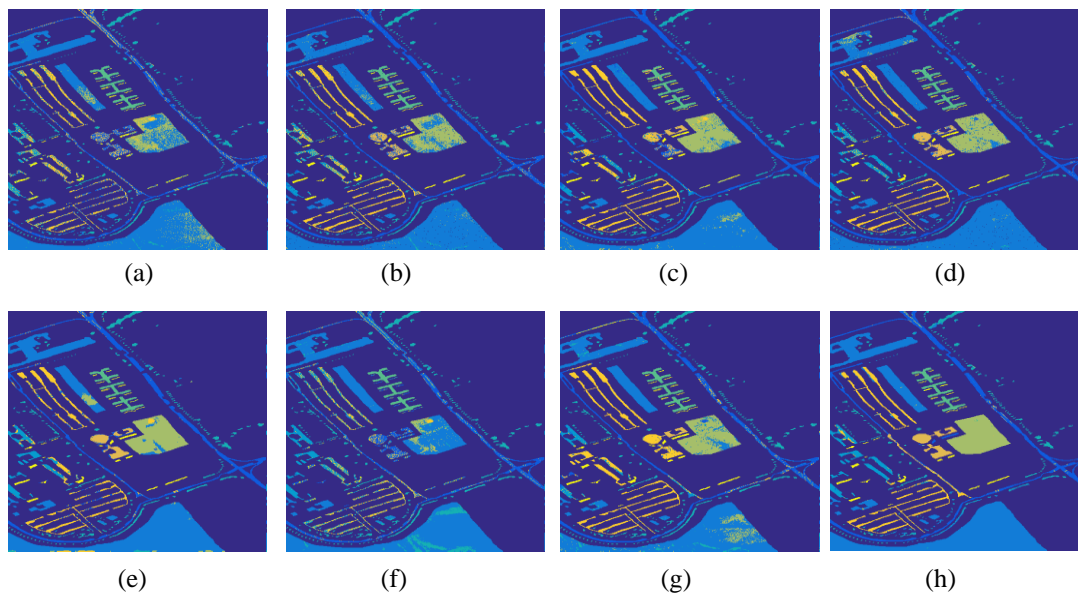


Fig. 4 Classification map of different classifiers on Indian Pines dataset. (a)-(h) without additional noise; (i)-(m) with additional noise (level of noise: $\sigma = 0.02$). (a) and (i)SMRL; (b) and (j)SVM; (c) SMLR-AP; (d) SVM-AP; (e) and (k)SVM-CK; (f) and (l)PCRC; (g) PCRC-AP; (h) and (m)LRR-PCRC.

TABLE 8. THE CLASSIFICATION RESULTS IN INDIAN PINES DATASET (1% TRAINING SAMPLES) (BEST RESULTS IN BOLD).

NO.	No additional noise							Noise level $\sigma = 0.02$						
	SMLR	SVM	SMLR-AP	SVM-AP	SVM-CK	PCRC	PCRC-AP	LRR-PCRC	SMLR	SVM	SVM-CK	PCRC	LRR-PCRC	
1	37.2±17.0	67.0± 18.2	92.1±6.6	93.3±4.9	67.0±21.2	21.5±28.5	90.1±3.4	97.8±2.9	13.7±7.5	37.4±22.2	74.3±17.0	0.9± 2.1	94.7±8.9	
2	61.4±7.1	63.3±8.3	70.1±4.5	66.9±5.9	70.4±11.1	63.6±11.4	68.6 ±4.3	80.7±5.0	34.4±5.0	39.5±5.1	64.0±8.8	30.5± 9.6	76.1±6.0	
3	25.3±6.8	45.9±10.7	70.6±6.3	78.9±6.5	64.9±16.2	35.7±15.5	77.3 ±7.1	67.3±5.1	10.3±3.2	20.6±6.4	48.9±10.1	1.1± 1.1	60.8±9.8	
4	20.6±9.7	52.3±12.7	55.1±12.7	59.6±14.4	54.9±21.9	24.5±14.3	53.6 ±10.4	83.9±10.3	2.6±2.8	26.7±12.4	61.5±12.9	0.3± 0.9	79.9±16.8	
5	47.6±13.4	72.8±17.0	75.4±13.0	74.8±13.1	71.5± 8.7	57.4±12.1	64.8 ±15.3	79.2±8.6	9.7±4.8	41.8±13.6	53.6±8.7	0.4± 1.0	76.0±7.4	
6	90.5±3.8	80.0±12.0	91.4±3.8	91.2±4.8	86.8± 8.1	95.3±1.7	84.7 ± 7.6	82.9±9.8	70.1±10.4	80.8± 6.8	85.4± 8.0	64.5± 12.1	80.7±10.9	
7	66.0±11.3	86.0±12.0	91.7±8.0	95.2±3.2	93.9± 5.8	31.7±14.0	99.1 ± 2.7	95.6±3.5	13.0±7.1	76.0±18.9	96.0±5.9	0± 0	93.4±2.2	
8	88.0±9.1	80.4±16.6	99.4±0.3	99.3±0.1	88.2± 8.4	98.7±2.1	99.6 ± 0.1	99.6±0.8	65.7±9.0	75.7±14.7	96.1±1.8	99.1± 0.5	99.9±0.1	
9	33.8±19.8	82.2± 24.9	58.3±19.8	97.2±8.7	86.1± 11.1	32.2±14.0	25.0 ±13.4	96.1± 8.3	0± 0	34.4±29.2	87.7± 5.7	0 ± 0	79.4±22.2	
10	34.5±8.6	57.6±13.7	60.3±15.5	59.9±11.1	60.0± 14.4	36.2±20.1	32.5 ± 12.7	78.3±5.0	15.5±4.6	31.9±12.6	53.1±6.0	3.7± 4.6	75.4±4.2	
11	69.0±5.5	72.7±4.7	86.9±5.8	76.0±7.7	79.3± 7.2	64.0±10.5	92.3 ± 4.2	92.1±2.9	58.9±4.9	74.4±7.8	82.9±5.3	88.5± 3.2	92.1±3.5	
12	28.6±8.4	44.8±7.0	48.5±14.6	48.8±12.0	46.3± 10.2	40.3±11.3	37.4 ± 9.2	77.9± 16.3	9.1±3.6	22.8± 12.4	45.8±8.3	1.5± 2.4	75.6±17.5	
13	93.6±5.7	87.7±10.1	98.8±0.6	98.8±0.4	93.5± 9.3	99.9± 0.2	99.5 ± 0	99.8± 0.2	46.1±8.9	90.1±6.1	91.2±5.3	40.2± 20.0	100.0±0	
14	89.4±5.5	86.6±8.6	92.4±6.7	87.5±7.8	87.3± 5.1	97.7±1.3	93.9 ± 5.1	99.6±0.4	93.6±3.7	90.3± 8.1	91.8±4.1	99.7± 0.1	99.6±0.4	
15	44.7±11.8	30.2±8.9	73.3±15.5	70.6±13.9	52.8±9.0	31.7±10.5	82.2 ± 6.7	90.1±10.9	16.1±3.8	19.3±7.2	63.9±11.6	4.1± 2.1	90.1±10.9	
16	82.5±7.8	84.3±2.4	87.9±6.1	97.4±2.6	93.9±7.6	96.2±2.1	96.1 ± 2.4	94.9±4.6	76.8±7.6	85.4±3.3	95.3±6.2	92.7± 3.4	95.2±4.8	
OA	61.2±2.1	67.2±1.2	78.6±1.8	75.7±2.0	73.6± 4.8	64.1± 2.2	76.5±1.8	86.4± 1.5	44.3± 1.5	56.2± 1.2	72.1± 1.8	49.5±1.5	84.4±1.3	
AA	57.1±2.3	68.4±2.3	78.3±1.5	81.0±1.2	74.8± 3.1	57.9± 2.5	74.8 ± 1.0	88.5± 1.6	33.5±1.8	52.9± 4.3	74.5 ± 2.1	32.9± 1.9	85.6±2.0	
k	55.2±2.4	62.6± 1.4	75.5±2.1	72.4±2.2	69.9± 5.5	58.8± 2.4	72.8 ± 2.1	84.5± 1.7	35.1± 1.8	49.2±1.6	67.9±2.1	38.9± 1.9	82.2± 1.6	



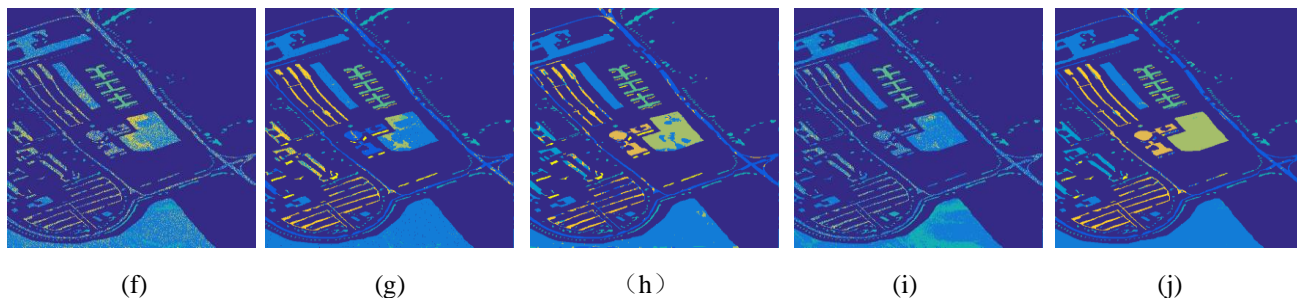


Fig. 5 Classification map of different classifiers on Pavia University dataset. (a)-(h) without additional noise; (i)-(m) with additional noise (level of noise: $\sigma = 0.02$). (a) and (i)SMRL; (b) and (j)SVM; (c) SMLR-AP; (d) SVM-AP; (e) and (k)SVM-CK; (f) and (l)PCRC; (g) PCRC-AP; (h) and (m)LRR-PCRC.

TABLE 9. THE CLASSIFICATION RESULTS IN PAVIA UNIVERSITY (1% TRAINING SAMPLES) (BEST RESULTS IN BOLD).

No.	No additional noise								Noise level $\sigma = 0.02$				
	SMLR	SVM	SMLR-AP	SVM-AP	SVM-CK	PCRC	PCRC-AP	LRR-PCRC	SMLR	SVM	SVM-CK	PCRC	LRR-PCRC
1	81.2±2.8	89.0±1.8	94.2±1.7	94.0±1.4	93.5±2.1	71.5±4.6	77.2±4.6	88.6±1.9	52.9±4.8	89.5±2.6	90.2±2.3	28.8±5.2	86.8±2.0
2	93.1±1.3	97.3±0.6	96.7±0.7	97.6±0.6	98.2±0.7	90.0±1.6	93.8±1.8	99.3±0.6	80.8±3.0	95.5±3.0	97.3±0.7	68.9±7.8	99.1±0.7
3	51.5±6.5	62.9±7.9	58.0±9.4	81.1±8.8	79.8±4.3	75.2±13.6	36.0±10.3	96.2±4.1	19.3±4.3	19.5±12.7	72.8±4.2	15.1±6.3	94.8±4.9
4	81.7±4.8	87.1±3.6	90.7±3.3	93.5±2.2	92.0±2.7	97.7±0.9	98.0±1.2	82.8±2.5	69.9±3.1	76.1±4.7	91.4±2.7	96.0±3.7	82.1±3.2
5	98.4±1.2	98.9±0.5	98.1±2.0	99.5±0.3	99.8±0.1	100.0±0	99.8±0.1	99.3±0.1	95.7±2.6	97.4±5.0	99.6±0.6	99.9±0	99.4±0.2
6	47.2±2.6	75.3±2.5	83.7±1.8	87.2±3.1	92.8±1.9	23.0±4.6	75.4±6.3	99.4±0.3	31.2±3.9	30.7±9.6	82.4±3.4	10.1±5.0	99.4±0.3
7	24.1±6.8	75.9±7.3	51.6±6.3	86.3±3.1	86.7±4.5	5.0±3.2	11.0±6.8	99.3±0.5	8.7±2.8	14.8±19.8	72.6±7.5	7.1±2.3	99.1±0.7
8	65.5±4.8	84.8±3.0	93.6±2.5	93.2±1.4	82.7±4.2	52.6±17.6	95.7±2.2	94.2±2.7	33.3±5.8	84.8±7.0	73.1±3.0	20.7±6.0	94.5±2.2
9	59.6±12.1	99.8±0.1	84.2±7.9	99.8±0.1	93.8±2.9	16.3±11.8	99.8±0.1	96.9±4.8	14.6±4.5	99.7±0.3	94.0±2.1	1.4±1.1	57.3±7.3
OA	77.9±0.7	89.4±0.4	90.6±0.4	94.1±0.5	93.8±0.4	71.9±0.4	84.4±1.1	95.8±0.4	59.5±1.0	78.6±0.8	90.0±0.4	48.5±3.7	94.5±0.4
AA	66.9±1.4	85.7±1.1	83.4±1.4	92.5±0.9	91.0±0.8	59.0±1.8	76.3±1.6	95.1±0.9	45.2±1.2	67.6±2.3	85.9±0.8	38.7±1.8	90.3±1.0
k	70.0±0.9	85.8±0.6	87.4±0.5	92.2±0.7	91.7±0.5	61.7±0.7	79.2±1.6	94.4±0.6	45.7±1.3	70.4±1.1	86.6±0.6	28.8±5.0	92.6±0.5

TABLE 10. THE TRAINING, VALIDATION AND TESTING RESULTS IN INDIAN PINES DATASET (5 SAMPLES PER CLASS FOR TRAINING, 5 SAMPLES PER CLASS FOR VALIDATION, REMAINING FOR TESTING)

NO.	SMLR			SVM			SVM-CK			PCRC			LRR-PCRC		
	Tr	Val	Ts	Tr	Val	Ts	Tr	Val	Ts	Tr	Val	Ts	Tr	Val	Ts
1	100±0	73.0±12.5	70.0±3.2	100±0	84.0±10.7	81.4±6.6	99.0±3.1	79.0±15.2	84.0±6.6	97.0±9.4	77.0±18.2	72.7±16.2	100±0	99.0±3.1	98.7±1.2
2	100±0	53.0±20.5	48.1±9.3	92.0±19.3	50.0±21.0	51.3±13.4	97.0±4.8	61.0±26.4	52.4±4.6	99.0±3.1	44.0±19.5	49.4±9.7	100±0	66.0±23.6	60.6±7.8
3	100±0	43.0±21.6	33.9±7.6	88.0±14.7	45.0±25.9	41.8±9.6	96.0±5.1	52.0±20.4	45.1±10.5	99.0±3.1	36.0±18.9	33.1±8.1	100±0	63.0±16.3	65.8±12.4
4	100±0	34.0±17.7	38.6±6.0	99.0±3.1	63.0±14.9	58.1±8.7	98.0±4.2	65.0±15.8	65.0±8.7	98.0±4.2	42.0±20.4	42.2±9.4	100±0	91.0±8.7	89.9±7.6
5	100±0	62.0±14.7	61.1±5.4	98.0±6.3	73.0±19.4	74.8±10.8	100±0	75.0±19.0	73.4±7.6	97.0±6.7	56.0±23.1	59.1±10.6	100±0	82.0±15.4	83.3±4.3
6	100±0	84.0±12.6	85.4±4.6	100±0	80.0±17.6	72.2±9.8	100±0	76.0±21.7	74.6±12.2	100±0	97.0±4.8	94.7±2.2	100±0	82.0±12.2	79.9±6.7
7	100±0	93.0±8.2	88.1±5.6	99.0±3.1	94.0±5.1	90.9±5.1	100±0	98.0±4.2	94.3±5.0	100±0	89.0±8.7	83.4±9.6	100±0	94.0±6.9	93.1±5.2
8	100±0	87.0±14.1	77.1±9.7	98.0±4.2	86.0±8.4	75.8±8.0	100±0	90.0±8.1	83.7±4.6	100±0	92.0±9.1	88.9±9.6	100±0	100±0	99.7±0.2
9	100±0	91.0±9.9	85.5±13.2	100±0	97.0±4.8	93.5±9.1	100±0	96.0±5.1	96.0±6.5	100±0	94.0±5.1	89.5±7.2	100±0	100±0	100±0
10	100±0	41.0±17.2	38.1±11.8	94.0±8.4	53.0±19.4	49.5±10.2	98.0±4.2	56.0±17.7	49.4±14.3	100±0	41.0±26.8	46.5±15.6	100±0	72.0±19.3	72.3±6.4
11	100±0	35.0±19.5	33.5±4.3	95.0±7.0	60.0±13.3	46.1±10.1	95.0±7.0	54.0±18.3	47.0±8.6	100±0	30.0±13.3	31.2±7.3	100±0	68.0±16.1	75.2±5.2
12	100±0	30.0±22.6	39.4±6.1	91.0±18.5	35.0±18.4	39.3±8.5	98.0±4.2	27.0±23.1	37.2±6.6	97.0±4.8	41.0±25.5	43.5±9.3	100±0	67.0±17.6	72.5±6.5
13	100±0	92.0±9.1	95.2±5.4	100±0	94.0±5.1	96.0±2.3	100±0	96.0±5.1	96.6±2.6	100±0	100±0	100±0	100±0	98.0±4.2	99.7±0.2
14	100±0	71.0±22.8	69.3±8.2	99.0±3.1	80.0±10.5	74.0±10.9	97.0±4.8	77.0±12.5	75.9±10.0	100±0	94.0±6.9	93.5±3.3	100±0	98.0±4.2	94.8±5.8
15	100±0	45.0±23.2	46.3±9.5	94.0±8.4	37.0±12.5	34.6±8.7	100±0	53.0±27.1	56.9±8.9	95.0±7.0	34.0±15.7	31.4±5.9	100±0	96.0±8.4	90.8±10.1
16	100±0	84.0±10.7	86.9±3.3	99.0±3.1	81.0±11.0	88.6±3.3	100±0	92.0±9.1	95.4±4.2	100±0	97.0±4.8	96.4±1.9	100±0	96.0±6.9	96.8±3.3
OA	100±0	63.6±3.2	50.5±2.9	96.6±4.6	69.5±3.6	56.0±3.4	98.6±1.4	71.6±5.0	58.8±2.9	98.8±1.2	66.5±3.3	55.0±1.8	100±0	85.7±3.6	78.0±2.1
AA	100±0	63.6±3.2	62.3±1.9	96.6±4.6	69.5±3.6	66.7±2.7	98.6±1.4	71.6±5.0	70.8±2.2	98.8±1.2	66.5±3.3	66.0±1.4	100±0	85.7±3.6	85.8±1.2
k	100±0	61.2±3.4	44.9±3.1	96.4±4.9	67.4±3.9	50.6±3.8	98.5±1.5	69.8±5.3	53.9±3.2	98.8±1.2	64.2±3.5	49.8±1.9	100±0	84.8±3.8	75.2±2.3

TABLE 11. THE TRAINING, VALIDATION AND TESTING RESULTS IN INDIAN PINES DATASET (5 SAMPLES PER CLASS FOR TRAINING, 5 SAMPLES PER CLASS FOR VALIDATION, REMAINING FOR TESTING)

No.	SMLR			SVM			SVM-CK			PCRC			LRR-PCRC		
	Tr	Val	Ts	Tr	Val	Ts	Tr	Val	Ts	Tr	Val	Ts	Tr	Val	Ts
1	96.0±6.9	94.8.0±21.4	37.7±9.0	82.0±13.9	66.0±18.9	55.1±3.1	94.0±13.4	66.0±20.6	60.5±8.6	76.0±12.6	33.0±20.0	21.7±7.3	99.0±3.1	72.0±15.4	59.0±5.1
2	99.0±3.1	62.0±26.9	57.9±10.7	83.0±22.6	62.0±19.8	54.8±14.5	97.0±6.7	68.0±19.3	65.8±12.5	90.0±9.4	69.0±26.0	69.4±12.9	100±0	80.0±18.2	75.6±10.6
3	99.0±3.1	53.0±13.3	61.1±5.4	68.0±27.4	43.0±23.5	48.3±15.5	93.0±15.6	55.0±25.0	64.6±12.0	95.0±5.2	71.0±15.2	78.7±9.2	98.0±4.2	97.0±4.8	94.6±3.9
4	100±0	86.0±11.7	88.8±5.9	98.0±4.2	82.0±18.1	82.6±11.0	100±0	76.0±17.7	79.1±10.9	100±0	94.0±8.4	95.3±3.4	100±0	73.0±8.2	77.9±9.4
5	100±0	100±0	98.9±0.1	100±0	98.0±6.3	96.8±4.0	100±0	100±0	98.6±2.7	100±0	100±0	99.9±0.1	100±0	100±0	99.2±0.3
6	100±0	45.0±19.0	45.6±5.5	91.0±12.8	43.0±22.1	53.0±12.0	98.0±6.3	49.0±21.8	60.2±9.1	78.0±9.1	24.0±12.6	30.4±9.0	100±0	97.0±4.8	94.0±3.5
7	99.0±3.1	170.0±13.3	359.6±11.2	99.0±3.1	94.0±10.7	91.3±2.4	99.0±3.1	87.0±12.5	86.6±5.8	95.0±7.0	68.0±15.4	59.7±10.2	100±0	99.0±3.1	99.4±0.4
8	100±0	40.0±15.6	48.6±6.7	81.0±17.2	63.0±25.8	64.3±12.2	95.0±8.4	65.0±20.6	68.1±8.4	90.0±10.5	33.0±15.6	38.2±9.9	98.0±4.2	81.0±12.8	77.6±13.7
9	96.0±5.1	85.0±9.7	85.6±8.1	100±0	99.0±3.1	99.8±0.1	100±0	91.0±11.0	91.8±6.3	68.0±18.7	49.0±23.3	45.5±17.1	100±0	67.0±22.6	73.8±9.8
OA	98.7±1.1	65.4±6.6	56.8±4.6	89.1±6.7	72.2±6.0	60.5±6.1	97.3±4.8	73±6.3	67.6±6.4	88.0±2.3	60.1±6.2	57.1±5.0	99.4±0.5	85.1±2.7	77.9±3.8
AA	98.7±1.1	65.4±6.6	64.9±1.8	89.1±6.7	72.2±6.0	71.8±1.7	97.3±4.8	73±6.3	75.0±2.8	88.0±2.3	60.1±6.2	59.9±1.5	99.4±0.5	85.1±2.7	83.5±2.3
k	98.6±1.2	61.1±7.4	46.7±4.6	87.7±7.6	68.7±6.7	51.5±5.9	97±5.4	69.6±7.0	59.4±7.2	86.5±2.6	55.1±7.0	46.0±4.7	99.3±0.6	83.2±3.0	72.3±4.2

TABLE 12. THE TIME CONSUMING COMPARISON IN INDIAN PINES AND PAVIA UNIVERSITY DATASET (20 TRAINING SAMPLES PER CLASS)

	20	SMLR	SVM	SMLR-AP	SVM-AP	SVM-CK	PCRC	PCRC-AP	LRR-PCRC
		Tr (s)	Ts(s)	AP:1.6	AP:1.6	-	-	AP:1.6	-
Indian Pines	20	-	-	AP:1.6	AP:1.6	-	-	AP:1.6	-
	Tr (s)	0.14	12.82	0.05	4.02	26.59	27.24	24.82	18.67
	Ts(s)	0.03	0.16	0.01	0.14	0.21	4.25	2.51	0.39
Pavia University	20	-	-	AP:4.3	AP:4.3	-	-	AP: 4.3	-
	Tr (s)	0.06	2.41	0.03	1.37	11.86	40.49	40.75	36.96
	Ts(s)	0.05	0.28	0.03	0.28	0.47	11.51	10.48	0.49

G. Extended Experiments and Analysis

In this section, in order to show the good performance of the proposed LRR-PCRC, we will conduct more experiments by comparing with the well-known method random forest (RF) [39] and its variations, including rotation forest (RoF) [40] and rotation random forest-kernel principal component analysis (RoRF-KPCA) [41].

TABLE 13. THE CLASSIFICATION RESULTS IN INDIAN PINES (20 TRAINING SAMPLES PER CLASS) (BEST RESULTS IN BOLD).

NO.	RoRF-KPCA					
	RF	RoF	Linear	RBF	Poly	LRRPCRC
1	84.81	75.29	89.63	92.22	91.30	99.7±0.9
2	32.57	46.55	58.65	57.03	61.93	86.7±5.8
3	42.01	50.31	51.97	51.43	51.43	80.5±4.8
4	53.12	66.96	83.12	85.17	82.82	96.5±3.2
5	76.66	76.79	86.38	84.85	85.01	89.6±4.5
6	67.95	75.25	91.34	92.36	92.01	96.3±2.9
7	98.08	78.33	99.23	99.23	99.23	96.9±3.9
8	80.57	78.83	92.70	92.68	93.62	99.8±0.1
9	90.00	78.00	98.00	99.00	98.00	99.0±3.1
10	48.61	58.39	76.03	76.17	73.85	84.7±4.4
11	44.04	37.71	53.16	54.57	52.07	89.2±2.4
12	42.56	47.86	80.70	79.06	81.95	91.2±6.0
13	93.73	92.97	99.06	98.87	98.73	99.8±0.2
14	80.00	81.03	89.10	89.46	91.11	99.5±0.9
15	46.08	44.08	62.87	62.47	61.40	98.9±1.1
16	96.42	95.33	97.05	97.05	98.32	97.8±1.9
OA	54.34	57.02	71.05	71.12	71.30	91.0±0.8
AA	67.33	67.73	81.81	81.98	82.04	94.1±0.5
k	49.06	51.99	67.61	67.65	67.85	89.7±0.9

The experiments are conducted under the situation that the training samples are 20 per class and remaining is used for testing. The kernel function of RoRF-KPCA has three different types, including lineal function, radial basis function (RBF) and polynomial (Poly) function. The results of RF, RoF and RoRF-KPCA are directly taken form [41]. Tables 13 and 14 show the classification results of the proposed LRRRCR, the RF and its variations in Indian Pines and Pavia University, respectively. As can be seen from these two tables, in general, the proposed method has obtained better classification accuracies than RF and its variations. In addition, the confusion matrices which are corresponding to the classification results of the proposed method in both datasets have been shown in Tables 15 and 16. We, hence, can conclude that proposed method has good performance again.

TABLE 14. THE CLASSIFICATION RESULTS IN PAVIA UNIVERSITY (20 TRAINING SAMPLES PER CLASS) (BEST RESULTS IN BOLD).

NO.	RoRF-KPCA					
	RF	RoF	Linear	RBF	Poly	LRRPCRC
1	65.56	72.45	79.01	75.72	76.05	75.8±3.7
2	99.99	99.98	99.97	99.97	99.99	90.9±3.3
3	99.15	99.56	99.37	99.12	99.33	98.4±1.6
4	65.81	81.88	78.27	77.69	77.43	88.1±8.2
5	82.25	90.67	89.88	89.93	91.95	99.5±0.2
6	56.11	70.80	72.91	72.90	74.75	98.1±2.6
7	54.02	60.30	58.33	66.88	68.08	99.6±0.3
8	68.35	68.96	70.55	70.72	71.12	92.9±2.3
9	84.51	82.83	90.89	88.18	90.72	98.1±3.4
OA	64.89	74.77	76.16	76.17	77.32	90.4±1.2
AA	75.08	80.83	82.13	82.35	83.27	93.5±0.6
k	56.61	68.31	69.79	69.79	71.16	87.5±1.5

TABLE 15. THE CONFUSION MATRIX OF PROPOSED LRR-PCRC IN INDIAN PAVIA (20 TRAINING SAMPLES PER CLASS)

		Predict Value																
		Category	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Actual Value	1	33.9	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0
	2	10.1	122.61	24.3	3.5	1.8	1.5	0	0	0	0	64.1	46.3	26.7	0	3.0	5	1.6
	3	0	17.9	655.5	65.8	2.9	0	0	0	0	5.1	13.8	5.0	46.2	0	0	1.8	0
	4	0	3.4	1.7	206.7	0	1.2	0	0	0	0	0	0	0.7	0	0	0.3	0
	5	2.4	0	3.6	0	427.4	0	31.9	0	0	0	0.9	1.4	0.7	2	0	6.7	0
	6	0	0.1	0.1	0	6.3	700.4	0	0	0	10.0	0	0.2	2.8	0	7.1	0	0
	7	0.2	0	0	0	0.2	0	12.6	0	0	0	0	0	0	0	0	0	0
	8	0	0	0	0	0	0	0	0	468.5	0	0	0	0	0	0.5	0	0
	9	0	0	0	0	0	0.0001	0	0	0.0099	0	0	0	0	0	0	0	0
	10	1.7	41.3	61.9	0	1	0	0	3.0	0	803.3	13.0	22.3	0.2	0.1	0.2	0	0
	11	0	170.1	6.8	1.3	4.5	12.2	0	11.8	3.5	46.1	2184.5	0.2	0	0	7	0	0
	12	0	3.8	18.8	10.7	0.6	0	0	0	0	9.3	0	542.2	0	0	0.4	8.2	0
	13	0	0	0.2	0	0	0	0	0	0	0	0	0	0	191.8	0	0	0
	14	0	0	0	0	0	1.6	0	2.9	0	0	0.3	0	0	1268.8	0.4	0	0
	15	0	0.4	0	0	0	0	0	0	0	0	0.1	2.7	0	0	356.3	0.5	0
	16	0	0	0	0	0	0	0	0	0	0	0	0	1.6	0	0	0	73.4

TABLE 16. THE CONFUSION MATRIX OF PROPOSED LRR-PCRC IN PAVIA UNIVERSITY (20 TRAINING SAMPLES PER CLASS)

		Predict Value									
		Category	1	2	3	4	5	6	7	8	9
Actual Value	1	5011.4	78.1	171.9	0.1	36.7	55.9	649.2	607.7	0	0
	2	1.9	16941	1.5	1154.3	0	130	0	400.7	0	0
	3	2	0.2	2046.4	0.6	0	0.8	1.5	27.5	0	0
	4	0.6	319.7	0.2	2683.8	6	33.1	0.1	0.5	0	0
	5	0.4	0.9	0	0	1318.4	0.4	0	4.9	0	0
	6	1.3	0.4	0	4.4	45.3	4915.7	6.1	35.8	0	0
	7	1.9	0	0.4	0	0	1.6	1305.6	0.5	0	0
	8	46.8	44.3	137.6	0	0	1.1	26.6	3405.6	0	0
	9	17.2	0	0	0	0.2	0	0.1	0	909.5	0

V. CONCLUSION

In this paper, a novel framework based on sparse constraint and prior information of HSIs, LRR-PCRC, has been proposed to extract efficient feature and classify HSIs. By imposing the sparse constraint to PCRC, the proposed RPCRC can tolerate environmental noise and thus extracts efficient features of HSIs and improves the classification accuracies. In addition, by adding the prior information of HSIs (ED and CI) to RPCRC, the proposed LRR-PCRC can improve the classification accuracies significantly. Experiments have been conducted to compare our proposed LRR-PCRC with other state-of-the-art methods, and the results show our methods have superior performances.

Our further work will be focused on time costs reduction. One promising way is to use mathematical models to find theoretical solutions; the other approach is seeking dimensionality reductions methods or semi-supervised learning

practically to further improve the classification accuracies and in the meantime reduce the computational time costs. With respect to the prerequisite of the proposed method, it just can be used for supervised learning which means we need to acquire some training samples for training the model, hence, it should be extended to unsupervised learning furtherly.

APPENDIX A

A. The solution of LRR-PCRC without ED nor CI

First, we introduce the auxiliary variable H to split the variable to simplify the problem of Eq. (12). Thus Eq. (12) can be converted to:

$$\hat{A} = \arg \min_A \{ \frac{1}{2} \| Y - XA \|_F^2 + \lambda \| H \|_1 + \beta / K \sum_{k=1}^K \| XA - X_{N_k} A_{N_k} \|_F^2 \}$$

$$s. t. H = A \tag{24}$$

Then, the corresponding augmented Lagrangian function for Eq. (24) can be rewritten as:

$$\min_{A,H,J} \left\{ \frac{1}{2} \|Y - XA\|_F^2 + \beta/K \sum_{k=1}^K \|XA - X_{N_k} A_{N_k}\|_F^2 + \lambda \|H\|_1 + \langle Y_1, H - A \rangle + \tau/2 \|H - A\|_F^2 \right\} \quad (25)$$

where $\langle Y_1, H - A \rangle = \text{trace}(Y_1^T (H - A))$, $\tau > 0$ is a penalty parameter. Y_1 is the Lagrange multipliers. Then the alternative optimization algorithm [27] can be applied to solve the model of Eq. (25). The details can be seen as follows.

Update H: fix A , then the H can be updated as:

$$H^{t+1} = \arg \min_H \left\{ \lambda/\tau^t \|H^t\|_1 + \frac{1}{2} \|H^t - A^t + Y_1^t/\tau^t\|_F^2 \right\} \quad (26)$$

The solution of Eq. (26) can be solved by the simple soft-threshold [34]:

$$\begin{aligned} H^{t+1} &= \text{soft}(A^t - Y_1^t/\tau^t, \lambda/\tau^t) \\ &= \max\{0, \text{abs}(e) - \lambda/\tau^t\} \times \text{sign}(e) \end{aligned} \quad (27)$$

where $\text{abs}(e)$ is the absolute value of e , $e = A^t - Y_1^t/\tau^t$ and sign is the sign function [35].

Update A: fix H , then the A can be updated as:

$$\begin{aligned} A^{t+1} &= \min_A \left\{ \frac{1}{2} \|Y - XA^t\|_F^2 + \beta/K \sum_{k=1}^K \|XA^t - X_{N_k} A_{N_k}^t\|_F^2 + \tau^t/2 \|H^t - A^t + Y_1^t/\tau^t\|_F^2 \right\} \end{aligned} \quad (28)$$

Then the solution of Eq. (28) can be achieved by the first-order derivation:

$$A^{t+1} = \left(X^T X + \beta/K \sum_{k=1}^K \bar{X}_{N_k}^T \bar{X}_{N_k} + \tau^t I \right)^{-1} (X^T Y + \tau^t H^t + Y_1^t) \quad (29)$$

B. The solution of LRR-PCRC without ED or CI

The corresponding augmented Lagrangian function for LRR-PCRC without ED or CI can be written as follows.

(1) LRR-PCRC without CI:

$$\begin{aligned} \min_{A,H,J} \left\{ \frac{1}{2} \|Y - XA\|_F^2 + \beta/K \sum_{k=1}^K \|XA - X_{N_k} A_{N_k}\|_F^2 \right. \\ \left. + \lambda \|\Gamma \odot H\|_1 + \langle Y_1, H - A \rangle + \tau/2 \|H - A\|_F^2 \right\} \end{aligned} \quad (30)$$

Update H: fix A , then the H can be updated as:

$$\begin{aligned} H^{t+1} &= \arg \min_H \left\{ \lambda/\tau^t \|\Gamma \odot H^t\|_1 + \right. \\ &\quad \left. \frac{1}{2} \|H^t - A^t + Y_1^t/\tau^t\|_F^2 \right\} \end{aligned} \quad (31)$$

The solution of Eq. (31) can be solved by the transformation of the soft-threshold rule:

$$\begin{aligned} H^{t+1} &= \text{soft}\left(A^t - \frac{Y_1^t}{\tau^t}, \Gamma \times \lambda/\tau^t\right) \\ &= \max\{0, \text{abs}(e) - \Gamma \times \lambda/\tau^t\} \times \text{sign}(e) \end{aligned} \quad (32)$$

where $e = A^t - Y_1^t/\tau^t$.

Update A: fix H , then the A can be updated as:

$$\begin{aligned} A^{t+1} &= \arg \min_A \left\{ \frac{1}{2} \|Y - XA^t\|_F^2 + \beta/K \sum_{k=1}^K \|XA^t - X_{N_k} A_{N_k}^t\|_F^2 + \tau^t/2 \|H^t - A^t + Y_1^t/\tau^t\|_F^2 \right\} \end{aligned} \quad (33)$$

Then the solution of Eq. (33) can be achieved by the first-order derivation:

$$\begin{aligned} A^{t+1} &= \left(X^T X + \beta/K \sum_{k=1}^K \bar{X}_{N_k}^T \bar{X}_{N_k} + \tau^t I \right)^{-1} \times \\ &\quad (X^T Y + \tau^t H^t + Y_1^t) \end{aligned} \quad (34)$$

(2) LRR-PCRC without ED:

$$\begin{aligned} \min_{A,H,J} \left\{ \frac{1}{2} \|Y - XA\|_F^2 + \beta/K \sum_{k=1}^K \|XA - X_{N_k} A_{N_k}\|_F^2 + \right. \\ \left. \gamma \|C \odot J\|_1 + \langle Y_2, J - A \rangle + \tau/2 \|J - A\|_F^2 \right\} \end{aligned} \quad (35)$$

where J is auxiliary variable to split the variable to let the model become more easily solvable, Y_2 is the Lagrange multipliers.

Update J: fix A , then the J can be updated as:

$$J^{t+1} = \arg \min_H \left\{ \gamma/\tau^t \|C \odot J\|_1 + \frac{1}{2} \|J^t - A^t + Y_2^t/\tau^t\|_F^2 \right\} \quad (36)$$

The solution of Eq. (36) can be solved by the transformation of the soft-threshold rule:

$$\begin{aligned} J^{t+1} &= \text{soft}\left(A^t - \frac{Y_2^t}{\tau^t}, C \times \lambda/\tau^t\right) \\ &= \max\{0, \text{abs}(e) - (\Gamma \times \lambda/\tau^t)\} \times \text{sign}(e) \end{aligned} \quad (37)$$

where $e = A^t - Y_2^t/\tau^t$.

Update A: fix J , then the A can be updated as:

$$\begin{aligned} A^{t+1} &= \arg \min_A \left\{ \frac{1}{2} \|Y - XA^t\|_F^2 + \beta/K \sum_{k=1}^K \|XA^t - X_{N_k} A_{N_k}^t\|_F^2 + \tau^t/2 \|J^t - A^t + Y_2^t/\tau^t\|_F^2 \right\} \end{aligned} \quad (38)$$

Then the solution of Eq. (38) can be achieved by the first-order derivation:

$$\begin{aligned} A^{t+1} &= \left(X^T X + \beta/K \sum_{k=1}^K \bar{X}_{N_k}^T \bar{X}_{N_k} + \tau^t I \right)^{-1} \times \\ &\quad (X^T Y + \tau^t J^t + Y_2^t) \end{aligned} \quad (39)$$

REFERENCES

- [1] J. Li, H. Zhang, Y. Huang, et al, "Hyperspectral image classification by nonlocal joint collaborative representation with a locally adaptive dictionary," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3707-3719, 2014.
- [2] J. Li, J. M. Bioucas-Dias, A. Plaza, "Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no.3, pp. 809-823, 2012.
- [3] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. 14, pp. 55-63, 1968.
- [4] T. Qiao, Z. Yang, J. Ren, et al, "Joint bilateral filtering and spectral similarity-based sparse representation: A generic framework for effective feature extraction and data classification in hyperspectral imaging," *Pattern Recognit.*, 2017.
- [5] F. Melgani, L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778-1790, 2004.

- [6] F. Cao, Z. Yang, J. Ren, et al, "Sparse Representation-Based Augmented Multinomial Logistic Extreme Learning Machine With Weighted Composite Features for Spectral-Spatial Classification of Hyperspectral Images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6263-6279, 2018.
- [7] F. Cao, Z. Yang, J. Ren, et al, "Local Block Multilayer Sparse Extreme Learning Machine for Effective Feature Extraction and Classification of Hyperspectral Images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5580-5594, 2019.
- [8] J. Li, J. M. Bioucas-Dias, A. Plaza, "Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4085-4098, 2010.
- [9] F. Cao, Z. Yang, J. Ren, et al, "Extreme sparse multinomial logistic regression: A fast and robust framework for hyperspectral image classification," *Remote Sens.*, vol. 9, no. 12, pp. 1255, 2017.
- [10] Q. Wang, Q. Li, and X. Li, "Hyperspectral band selection via adaptive subspace partition strategy," *IEEE J. Sel. Top. App. Earth Obs. and Remote Sens.* vol. 12, no. 12, pp. 4940-4950, 2019.
- [11] Q. Wang, F. Zhang, and X. Li, "Optimal clustering framework for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.* vol. 56, no. 10, pp. 5910-5922, 2018.
- [12] X. Xu, J. Li, S. Li, and A. Plaza, "Subpixel Component Analysis for Hyperspectral Image Classification," *IEEE Trans. Geosci. Remote Sens.* vol. 57, no. 8, pp. 5564-5579, 2019.
- [13] Q. Wang, S. Liu, J. Chanussot, and X. Li, "Scene classification with recurrent attention of VHR remote sensing images," *IEEE Trans. Geosci. Remote Sens.* vol. 57, no. 2, pp. 1155-1167, 2018.
- [14] Q. Hao, S. Li, and X. Kang, "Multilabel Sample Augmentation-Based Hyperspectral Image Classification," *IEEE Trans. Geosci. Remote Sens.* (2020).
- [15] C. Liu, J. Li, and L. He, "Superpixel-based semisupervised active learning for hyperspectral image classification," *IEEE J. Sel. Top. App. Earth Obs. and Remote Sens.* vol. 12, no. 1, pp. 357-370, 2018.
- [16] J. Wright, Y. Ma, J. Mairal, et al, "Sparse representation for computer vision and pattern recognition." *Proc. IEEE*, vol. 98, no. 6, pp. 1031-1044, 2010.
- [17] Y. Chen, N. M. Nasrabadi, T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973-3985, 2011.
- [18] J. Liu, Z. Wu, J. Li, et al, "Probabilistic-Kernel Collaborative Representation for Spatial-Spectral Hyperspectral Image Classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 2371-2384, 2016.
- [19] L. Zhang, M. Yang, X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" *Computer vision (ICCV), 2011 IEEE international conference on. IEEE*, pp. 471-478, 2011.
- [20] M. Yang, L. Zhang, D. Zhang, et al, "Relaxed collaborative representation for pattern classification," *IEEE Conference on Computer Vision and Pattern Recognition, IEEE*, pp. 2224-2231, 2012.
- [21] J. Waqas, Z. Yi, L. Zhang, "Collaborative neighbor representation based classification using l_2 -minimization approach," *Pattern Recognit. Lett.*, vol. 34, no. 2, pp. 201-208, 2013.
- [22] S. Cai, L. Zhang, W. Zuo, et al, "A probabilistic collaborative representation based approach for pattern classification," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2950-2959, 2016.
- [23] W. Li, Q. Du, "Joint within-class collaborative representation for hyperspectral image classification," *IEEE J. Sel. Top. App. Earth Obs. and Remote Sens.*, vol. 7, no. 6, pp. 2200-2208, 2014.
- [24] J. Li, H. Zhang, Y. Huang, et al, "Hyperspectral image classification by nonlocal joint collaborative representation with a locally adaptive dictionary," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3707-3719, 2014.
- [25] J. Li, H. Zhang, L. Zhang, et al, "Joint collaborative representation with multitask learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5923-5936, 2014.
- [26] T. Qiao, et al, "Effective denoising and classification of hyperspectral images using curvelet transform and singular spectrum analysis." *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 119-133, 2017.
- [27] X. Lu, Y. Wang, Y. Yuan, "Graph-regularized low-rank representation for destriping of hyperspectral images." *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 4009-4018, 2013.
- [28] G. Chen, S. Qian, "Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 973-980, 2011.
- [29] Q. Yuan, L. Zhang, H. Shen, "Hyperspectral image denoising employing a spectral-spatial adaptive total variation model," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3660-3677, 2012.
- [30] L. Zhang, et al, "On the dimensionality reduction for sparse representation based face recognition," *2010 20th International Conference on Pattern Recognition, IEEE*, 2010.
- [31] J. Wright, A. Y. Yang, A. Ganesh, et al, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210-227, 2009.
- [32] Y. Chen, N. M. Nasrabadi, T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.* vol. 49, no. 10, pp. 3973-3985, 2011.
- [33] Q. Wang, X. He and X. Li, "Locality and Structure Regularized Low Rank Representation for Hyperspectral Image Classification," *IEEE Trans. Geosci. Remote Sens.* vol. 57, no. 2, pp. 911-923, 2019.
- [34] M. Afonso, J. Bioucas-Dias, M. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2345-2356, 2010.
- [35] J. Li, J. M. Bioucas-Dias, A. Plaza, "Hyperspectral image segmentation using a new Bayesian approach with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3947-3960, 2011.
- [36] Y. Zhou, J. Peng, C. L. P. Chen, "Extreme learning machine with composite kernels for hyperspectral image classification," *IEEE J. Sel. Top. App. Earth Obs. and Remote Sens.* vol. 8, no. 6, pp. 2351-2360, 2015.
- [37] <https://www.fst.um.edu.mo/en/staff/fstycz.html>
- [38] <http://www.lx.it.pt/~jun/>
- [39] Breiman, Leo, "Random forests," *Mach. Learn.* vol. 45, no. 1, pp. 5-32, 2001.
- [40] J. J. Rodriguez, L. I. Kuncheva, and C. J. Alonso. "Rotation forest: A new classifier ensemble method," *IEEE Trans. Pattern Anal. Mach. Intell.* vol. 28, no. 10, pp. 1619-1630, 2006.
- [41] J. Xia, N. Falco, J. A. Benediktsson, P. Du, and J. Chanussot, "Hyperspectral image classification with rotation random forest via KPCA," *IEEE J. Sel. Top. Appl. Earth Observ.* vol. 10, no. 4, pp. 1601-1609, 2017.