

# Hyperspectral Image Classification via Joint Sparse representation of Multi-layer Superpixels

Haifeng Sima<sup>1</sup>, Aizhong Mi<sup>1</sup>, Xue Han<sup>1</sup>, Shouheng Du<sup>1</sup>, Zhiheng Wang<sup>1</sup> and Jianfang Wang<sup>1</sup>

<sup>1</sup>Department of Computer science and technology,  
Henan Polytechnic University, Jiaozuo [e-mail: smhf@hpu.edu.cn]  
[e-mail: wzhenry@eyou.com]

\*Corresponding author: Zhiheng Wang

*Received November 24, 2017; revised January 31, 2018; revised March 8, 2018; accepted March 20, 2018;  
published October 31, 2018*

---

## Abstract

In this paper, a novel spectral-spatial joint sparse representation algorithm for hyperspectral image classification is proposed based on multi-layer superpixels in various scales. Superpixels of various scales can provide complete yet redundant correlated information of the class attribute for test pixels. Therefore, we design a joint sparse model for a test pixel by sampling similar pixels from its corresponding superpixels combinations. Firstly, multi-layer superpixels are extracted on the false color image of the HSI data by principal components analysis model. Secondly, a group of discriminative sampling pixels are exploited as reconstruction matrix of test pixel which can be jointly represented by the structured dictionary and recovered sparse coefficients. Thirdly, the orthogonal matching pursuit strategy is employed for estimating sparse vector for the test pixel. In each iteration, the approximation can be computed from the dictionary and corresponding sparse vector. Finally, the class label of test pixel can be directly determined with minimum reconstruction error between the reconstruction matrix and its approximation. The advantages of this algorithm lie in the development of complete neighborhood and homogeneous pixels to share a common sparsity pattern, and it is able to achieve more flexible joint sparse coding of spectral-spatial information. Experimental results on three real hyperspectral datasets show that the proposed joint sparse model can achieve better performance than a series of excellent sparse classification methods and superpixels-based classification methods.

---

**Keywords:** Hyperspectral image classification, Multi-layer superpixels, Joint sparse representation, Discriminative optimization sampling, Reconstruction matrix

---

This work was supported by the NSFC (61572173,61602157), the Science and Technology Planning Project of Henan Province (162102210062), the Key Scientific Research Fund of Henan Provincial Education Department for higher school (15A520072), Doctoral Foundation (B2016-37) and Young Scholar sponsored of Henan Polytechnic University, and Henan Postdoctoral. Henan Science and Technology Innovation Outstanding Youth Program (184100510009)

## 1. Introduction

**H**yperspectral imaging has been widely used in remote sensing information collection, which is acquired through hundreds of narrow continuous spectral bands to constitute the data set. In hyperspectral image (HSI), each pixel is a high dimensional mixture vector consisting of different spectral response that result multiple channels and complex texture for difficult classification[1].

The task of hyperspectral image classification is to assign the image pixels in hyperspectral images to a predefined class label. The existing classification methods can be divided into two categories according to whether or not to adopt training samples, namely, supervised and unsupervised classification. Supervised classification requires a labeled sample set for training the classifier to identify unlabeled pixels. The unsupervised classification is also called clustering technology, makes use of the information contained in unlabeled data to classify HSI pixels without any training samples. In general, the supervised methods can improve the accuracy and ensure the efficiency compared to unsupervised methods.

In the past decades, HSI classification has been a very active research topic in remote sensing. Many classification methods have been developed in order to improve the classification accuracy and cut the dependence on the labeled samples. Among them, classification methods based on support vector machines [2,3,4] and multinomial polynomial regression(MLR)[5,6] have been extensively investigated and they have shown powerful capability of classification. Aiming at the high dimension and less samples, to compress raw data into low dimensional space or subspace learning are also research hotspots to reduce the difficulties of classification. A number of excellent dimensionality reduction methods have been exploited for HSI classification, such as manifold learning[7], local fisher discriminant analysis [8] and multiple kernels[9].In addition, the classification strategy based on ensemble classifier and random forest provides a structure analysis method for HSI classification [10,11].

Although the aforementioned methods have improved the classification efficiency through spectral information, their results often appear noisy in many cases. This is mainly due to the blindness of feature extraction and representation. In addition, the training samples tend to be very limited, and therefore it is difficult to extract the essential features of each class for more accurate classification. In order to further improve the classification performance, more and more new researches attempt to explore spectral-spatial knowledge of pixels. It is based on the assumption that the pixels in adjacent regions should have similar spectral properties, which consist of the same materials.

In recent years, with the maturity of the compressed sensing theory, sparse representation (SR) has gradually been embedded into the computer vision and image processing tasks and applications, such as face recognition[12], image restoration[13], target detection[14], humane action recognition[15] and hyperspectral unmixing[16]. The sparse representation has become an important technology to solve the visual cognitive tasks, and more related researches based on sparse representation in HSI classification appears [17-25]. On the one hand, many superpixels-based methods effectively provide the context information of local regions for pixels classification [26,27].

To the classification task, feature extraction and representation of pixel characteristics is the key to accurate classification [28, 29]. In particular, SR can represent the linear components of

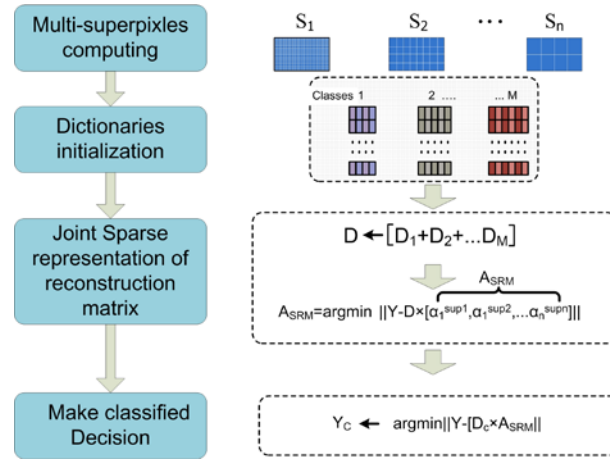
high dimensional signals based on a number of non-zero coefficients and pre-defined dictionaries, which demonstrates that SR has strong ability in classification.

Chen[17] proposed a classification method integrating the contextual information of adjacent area into a sparse representation model called Simultaneous Subspace Pursuit, which performs better than the classical supervised classifier support vector machines in most cases. Later, the sparse representation is further combined with the kernel space optimization method to obtain new progress in[18]. Zhang[19] et al. proposed a non-local weighted joint sparse representation model to improve the classification performance in the exploration of sparse solver. In[20], a correntropy-based metric has been proposed for robust classification for handling non-Gaussian noise and large outliers. Fang et al. believed that different scale regions can be incorporated into the complementary information for classification, and they proposed a multi-scale adaptive sparse representation classification model of HIS[21]. On one hand, superpixels-based methods are proposed to effectively exploit the spatial information of the HSI, but they are all work in single scale by experience [22, 23]. The recovered sparse coefficients are utilized to determine the class label of the superpixel. On the other hand, the researchers also pay attention to the introduction of dimensionality reduction. In[24], a sparse matrix transform-based LDA method is proposed for HSI classification. The sparse matrix can be used to decompose the high dimension matrix by sparse parameterization. In this way, the LDA model can estimate optimal discriminant vector more accurately even in the case of limited training samples. Zhai et al. proposed an weighted sparse graph embedding method for HSI classification [25]. It offers data-adaptive neighborhoods as training pixels for test pixels to implement sparse coding, which is more robust to noise. Based on SR and manifold learning, Huang et al. proposed a sparse discriminant embedding method preserves the sparse reconstructive relations and explicitly boosts the discriminating information from inter-manifold structure of training samples[30].

It is obviously that the spatial context information is supremely helpful for the feature representation and the extraction of pixels. Therefore, most aforementioned researches are committed to utilize similar and adjacent regions context information to achieve sparse modeling for classification. However, the existing models still suffer from following problems: 1) the strategy of region selection for context information is rather blindness in local area; 2) the utilization of context information is monotonous and performs poorly when faced with complex and changing materials; 3) the sparse reconstruction model and context information are lack of completeness. In order to solve the above problems, we believe that the multiple superpixels can provide complete context information for a specific pixel, and the combined superpixel regions can describe the class features better and improve the classification accuracy. Motivated by the above consideration, we propose a HSI classification method based on the joint sparse representation based on group sparsity of discriminative optimal sampling on multi-layer superpixels in this paper. We utilize different types of segmentation method to extract multiple-scales superpixels, and each test pixel is contained in a set of superpixels from different segmentations. Thus, a group of represent pixels are chosen from multiple superpixels as similar reconstruction matrix(SRM) of the test pixel the group feature of and its group sparse coefficients can be estimated by orthogonal matching pursuit algorithm. Finally, the class label of the test pixel is determined by the minimum reconstruction error of SRM computed by group sparse coefficients and subdictionary of each class.

The rest of this paper is organized as follows: the related works are reviewed in Section 2. In Section 3, the Multi-layer superpixels-based joint sparse representation classification (MSJSRC) is introduced in detail. Several experiments were conducted to show the

effectiveness of the proposed method as reported in Section 4. Furthermore, the comparisons with several state-of-the-art HSI classification methods are given. Finally, conclusions are presented in Section 5. The proposed MSJSRC framework is illustrated in Fig. 1.



**Fig. 1.** The Multi-layer superpixels-based joint sparse representation classification framework for hyperspectral image classification

## 2. Related Work

The research in this paper is based on the superpixels extraction, sparse representation classification and joint sparse representation. Hence, we introduce these technologies that will be adopted in this paper.

### 2.1 Superpixels extraction

In many superpixels extraction methods, mean shift segmentation is a pixel clustering process based on the maximums estimation of probability density in the spatial and feature space, which can smooth homogeneous area and obtaining over-segmentation results with complete shape on large scales[32]. Another excellent method also perform well at local detail of complex shape or relative dispersion called graph-cut, and its basic principle is to map the image to an undirected graph and utilize the minimum cut theory to divide the pixels as segmentation[31]. In addition, another method is to construct a relatively complete homogeneous region with the linear clustering strategy of seed expansion named SLIC [33]. These three methods adopt different strategies of superpixel calculation, who are helpful for building complete dictionary of pixel representation. Therefore, we choose the three above-mentioned methods to generate a complete superpixels set to construct the complete sparse representation. The following are brief reviews on the principles of the three methods.

#### 2.1.1 Mean-Shift

The mean shift image segmentation is a straightforward extension of the discontinuity preserving smoothing by probability density estimation. Each pixel is associated with a significant mode of the joint domain density located in its neighborhood, after nearby modes were estimated by probability density estimation with feature space analysis technique. One can get various segmentation superpixels results with different employed kernel bandwidths of corresponding feature and spatial resolution.

Let  $x_i$  and  $z_i$ ,  $i=1,\dots,n$  be the d-dimensional input and filtered image pixels in the joint spatial-range domain and  $L_i$  is the label of the  $i$ -th pixel in the segmented image.

- 1) Compute the mean shift vector  $M(x)_i$  iteratively until converged, and store all the information about the d-dimensional convergence point in  $z_i$
- 2) Delineate in the joint domain the clusters  $\{C_p\}$   $p=1\dots m$  by grouping together all  $z_i$  which in the  $h_s$  and  $h_r$  range domain
- 3) For each  $i=1,\dots,n$ , assign  $L_i=\{p \mid z_i \in C_p\}$
- 4) Optional: Merge meaningless regions containing less than given pixels.

### 2.1.2 Graph cut-NNG

Graph-based image segmentation techniques generally represent the problem in terms of a graph  $G = (V, E)$  where each node  $v_i \in V$  corresponds to a pixel in the image, and the edges in  $E$  connect certain pairs of neighboring pixels. The weights of all edges are assigned according to the similarity of some property, and the nodes in the graph are clustered by finding minimum cuts in the graph, which making data within the same group as similar as possible and within the different group as different as possible. The generated superpixels are the minimum spanning tree of the image. The internal difference of a component  $C$  is defined as the largest weight in the minimum spanning tree:

$$Int(C) = \max_{e \in \text{mst}(C,E)} \omega(e) \quad (1)$$

The differences between two components  $C_1$  and  $C_2$  is defined as the minimum weight edge connecting the two components:

$$Dif(C_1, C_2) = \min_{v^i \in C_1, v^j \in C_2, (v^i, v^j) \in E} \omega(v^i, v^j) \quad (2)$$

By means of the definition of the internal difference and the difference between two components, as well as comparison with in their relationship, we can get the judgment conditions of merging:

$$D(C_1, C_2) = \begin{cases} true & \text{if } Dif(C_1, C_2) > MInt(C_1, C_2) \\ false & \text{otherwise} \end{cases} \quad (3)$$

$MInt$  is the minimum internal difference,

$$MInt(C_1, C_2) = \min(Int(C_1) + \tau C_1, Int(C_2) + \tau C_2) \quad (4)$$

The threshold function  $\tau$  controls the degree to which the difference between two components must be greater than their internal differences:  $\tau=k/|C|$ , where  $k$  is a scale parameter and a larger  $k$  causes a preference for larger component.

### 2.1.3 SLIC

SLIC clustering pixels based on their property similarity and neighbour pixels in the image plane, which can generate compact and even superpixels effectively. SLIC predefined a desired superpixels number  $K$  in equally sized approximately. When the size of image is  $M \times N$  pixels, each superpixel contains about  $M \times N / K$  pixels. The  $K$  superpixel cluster centers should be chosen at every grid interval  $S = \sqrt{M \times N / K}$  by sampling  $K$  seed locations corresponding to the lowest gradient position. Since the spatial approximate size of a superpixel is approximately  $S^2$ , SLIC implements the clustering process pixels lay within a  $2S \times 2S$  area around the superpixel center on image plane. This becomes the search area for the pixels nearest to each cluster center.

- 1) Sampling the lowest gradient position at regular grid in an  $S \times S$  neighborhood as cluster centers.
- 2) **for** each cluster center  $C_k$  **do**
- 3) Gather the best matching pixels from a  $2S \times 2S$  square neighborhood around the cluster center according to the distance measure in  $[l, a, b, x, y]$  feature space.
- 4) **end for**
- 5) Compute new cluster centers and residual error  $E$ , until  $E \leq \text{threshold}$ .
- 6) Enforce connectivity by relabeling disjoint segments with the labels of the largest neighboring cluster.

## 2.2 Sparse representation and Classification on HSI

In the sparsity representation model of computer vision, it is assumed that a test pixel can be reconstructed by a sparse linear combination of some atoms from a dictionary [17]. Let  $y \in \mathbb{R}^{d \times 1}$  be a test pixel, where  $d$  is the feature vectors dimension of training image. A given matrix  $D \in \mathbb{R}^{M \times N}$  consist of some atoms as dictionary, the test pixel can be approximately represented by multiplying the dictionary with a sparse vector  $\alpha$ . The test pixel  $y$  can be sparsely expressed. The sparse vector can be recovered by solving the following optimization problem:

$$y = D\alpha \quad (5)$$

The sparse vector can be recovered by solving the following optimization problem:

$$\alpha = \arg \min \|y - D\alpha\|_F \quad \text{subject to} \quad \|\alpha\|_0 < K_0 \quad (6)$$

In HSI classification, supposing the datasets have  $C$  distinct classes and stack a few training pixels from the  $i$ -th class as columns of a dictionary  $D_c = [D_1^{n_1}, D_2^{n_2}, D_3^{n_3}] \in \mathbb{R}^{M \times \sum_1^n n_c}$ , Where the  $D_c$  is a low-dimension space included  $\sum_1^n n_c$  training samples, and  $M$  refers to the number of bands of the HSI. The hyperspectral pixel belonging to the  $i$ -th class can be compactly represented as a linear combination of the given training samples. With the optimal solution of sparse vector  $\alpha$ , the class label of  $y$  is decided based on the following criterion of minimum reconstruction error:

$$c = \arg \min \|y - D_c^{n_c} \alpha\|_F \quad c = \{1, 2, \dots, C\} \quad (7)$$

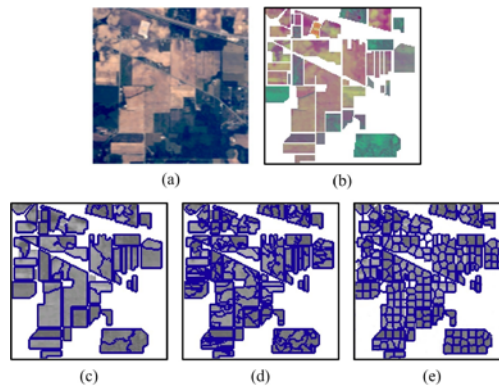
## 3. Multi-layer Superpixels-based joint sparse representation classification (MSJSRC)

In order to achieve complete and optimized sparse representation, we present a multi-layer superpixels supervised sparse representation classification method in this section. According to the aforementioned analysis, the combination of adjacent pixels for sparse representation can produce effective encoding of class characteristics and be useful for improving classification accuracy. In addition, multi-layer superpixels as grouping cues allows us to effectively encode complex image structures in joint sparse model [39]. The implementation detail of the method is described below.

### 3.1 Adaptive Superpixels extraction

Existing superpixel segmentation algorithms are always running on the gray scale and the three component image. These methods are not suitable for conducting on high-dimensional hyperspectral image directly that lead to large computing load. Considering the high correlation of hyper-spectral features and information redundancy, dimensionality reduction is an ideal solution to map high-dimensional data to low dimensional space through a linear projection and retain more of the original data at the same time. PCA[40] is able to maximize the intrinsic information of the data after dimension reduction by measuring the size of the variance of the data in the projection direction. Therefore we adopt PCA method to extract first three principle components of hyper-spectral data to reconstruct three component image for superpixels extraction.

We set up their parameters for extracting various scales of superpixels according to their respective advantages of the three methods. The main parameters of mean-shift are spatial and spectral feature bandwidth and minimum area:  $hs$ ,  $hr$ , and  $Min-area$ . They are set rely on the image size and roughness of the input image and get the segmentation  $Seg_{MS}$ :  $hs = \sqrt{M \times N / K \times K}$ ,  $hr=16$ , and  $Min-area = M \times N / K \times K$ . Where  $M$  and  $N$  are the size of the input image and  $K$  is the classes number containing in the image. The gauss smooth  $\delta = 0.8$  constant scale parameter  $k = 100$  and for medium scale super pixels  $Seg_{NNG}$ . The superpixels number of SLIC tend to be smaller to obtain complete homogeneous areas, thus the number  $K = M \times N / 80$  is proper for getting ideal homogeneous superpixels  $Seg_{SLIC}$  by experience. The superpixels are show in Fig. 2.



**Fig. 2.** Discriminative optimal sampling of Multi-layer superpixels for joint sparse representation classification. (a) Input image (b) false color image (c)MS.(d)NNG. (e)SLIC.

### 3.2 Multi-layer superpixels joint sparse representation classification

for joint sparse representation classification After superpixels extraction, any pixel  $\mathbf{y}$  of the input data corresponding to a superpixel set, and these superpxiels are considered having similar materials with  $\mathbf{y}$  and multiple scales superpxiels will provide complete class information corresponding to pixel  $\mathbf{y}$ . Thus, the combined superpixels set around  $\mathbf{y}$  can be reconstructed by the combined representation of multiple scales superpxiels as

$$Y = \begin{bmatrix} Y_{sup_1} \\ Y_{sup_2} \\ \dots \\ Y_{sup_L} \end{bmatrix} = \begin{bmatrix} D\alpha_{sup_1} \\ D\alpha_{sup_2} \\ \dots \\ D\alpha_{sup_L} \end{bmatrix} = D \begin{bmatrix} \alpha_{sup_1} \\ \alpha_{sup_2} \\ \dots \\ \alpha_{sup_L} \end{bmatrix} = D \{A\}_I^L \quad (8)$$

where  $L$  is the superpixels number,  $A_{sup}$  is the sparse coefficients matrix corresponding to  $\mathbf{y}$  which consists of the indexes of the selected atoms in the joint dictionary  $D_{sup}$ . The proposed model enables pixels from all corresponding superpixels  $[sup_1, sup_2, \dots, sup_L]$  to be represented by a few common atoms and nonzero rows of the sparse coefficients matrix. Thus the recovering of  $[sup_1, sup_2, \dots, sup_L]$  can be solved by the following joint representation model:

$$\{A\}_I^L = \arg \min \sum_{k=1}^L \|\mathbf{y} - D A_{sup k}\|_F \quad \text{subject to} \quad \|\{A\}_I^L\|_{row,0} < K_0 \quad (9)$$

where  $\{A\}_I^L = [A_1, A_2, \dots, A_L]$  is the joint sparse coefficients of corresponding matrix of test pixel  $\mathbf{y}$ , and  $\|\cdot\|$  is the Frobenius norm.  $K_0$  is the upper bound of the sparsity level that is the maximum number of selected atoms from the dictionary for the best simultaneous approximation.

However, the estimation of sparse feature vectors of a test pixel using the reconstruction of the combined superpixels region exist problems below: 1. For all composite superpixels regions of test pixel  $\mathbf{x}$  contain redundancy information, and the sparse pixel estimation results in the overlapping regions is not unique for classification of test pixel. 2. The reconstruction process of multiple combination superpixels costs more computational complexity.

Therefore, we proposed a discriminative optimal sampling pursuit algorithm for approximation sparse coding for two reasons above. The joint sparse of signals share a common sparse component and discriminative information can improve the promising recognition ability[34,35,36]. The schematic of the sparse coding of our method is shown in Fig. 3.

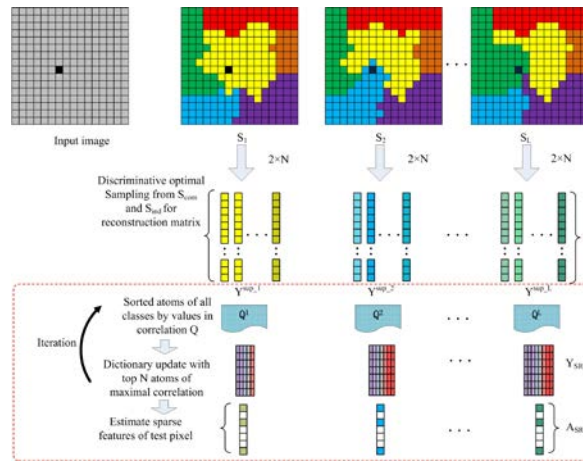


Fig. 3. The illustration of the sparse coefficients estimation with multi-layer superpixels.

Firstly, a joint dictionary  $D=[D_1 \dots D_m \dots D_M] \in R^{B \times W}$  is constructed by selected labeled training samples with  $M$  classes and  $W$  is the atoms number, and  $B$  denotes the number of spectral bands. Suppose a test pixel  $\mathbf{y}=[y_1, y_2, \dots, y_B]$  with  $B$  bands, find all corresponding superpixels in all segmentation superpixels. All pixels in each superpixel are initialized as matrix  $Y^{sup-k}$ , the correlation between  $\mathbf{y}$  and  $Y^{sup-k}(B \times P)$  is calculated and result as a column vector  $Q_p^k = \mathbf{y} \cdot Y^{sup-k}$ . In order to reduce the computational complexity and redundancy information, we decomposed the pixels set  $S^{sup-k} \{P_1, P_2 \dots P_p\}$  into two parts according to the



eigenvalues in the correlation vector  $Q_P^k$ . All the atoms eigenvalues are sorted in descending order in a sequence  $S^{sup-k}$ . The first half part  $S_{com}^{sup-k} = \{P_1, P_2, \dots, P_{P/2}\}$  contained more common informative components and the other part  $S_{ind}^{sup-k} = \{P_1, P_2, \dots, P_{P/2}\}$  contained more independent informative components. The homogeneous attributes of one class is easy to be described with common part, and independent informative components express expansibility and tolerance of the corresponding class. According to the sorted correlation degree in  $S^{sup-k}$ , the first  $N$  pixels are selected from  $S_{com}^{sup-k}$  and  $S_{ind}^{sup-k}$  to form the similar reconstruction matrix  $Y_{SRM}$  of the test pixel  $y$ . Thus, a novel discriminative sampling matrix norm is designed as substitution of combined superpixels set:

$$A_{SRM} = \arg \min \|Y_{SRM} - DA_{SRM}\|_F \quad \text{subject to} \quad \|A_{SRM}\|_{row,0} < K_0 \quad (10)$$

The joint sparse represent method fuse similar reconstruction matrix information for accurate classification, and the detailed class-labeled OMP algorithm is shown in **Algorithm 1**. The approximate solution of the similar matrix can be solved by the Orthogonal Matching Pursuit algorithms. The correlation degree between  $Y_{SRM}$  and  $D_I$  is defined as  $R_{i=1}^x = D_{i,i}^T Y_{SRM}$ ,  $D_{i,i=0} = D$ . In each iteration, we choose the best represent atoms from  $D$  to update a segmental optimal index dataset  $I$  and the  $D_I$  is utilised to estimate the joint sparse coefficients  $A_{SRM}$  of the test pixel  $y$ . The sparsity-constrained optimization problem of  $A_{SRM}$  in (10) is obtained by the least squares method [42] and the analytic solution is computed by

$$A_{SRM} = (D_I^T D_I)^{-1} D_I^T Y_{SRM} \quad (11)$$

With the recovering of sparse representation matrix  $A_{SRM}$ , the label of the test pixel  $x_i$  can be decided by the lowest total representation error:

$$C_i = \arg \min_c \|Y_{SRM} - D_C A_{SRM}\|_F \quad (12)$$

**Algorithm 1. The pseudo-codes of the Multi-layer superpixels joint sparse representation classification**

**Input:** A HSI data matrix set  $\mathbf{X} = [x_1, x_2, x_3, \dots, x_N] \in \mathbb{R}^{B \times V}$ , mutiple superpixels [Seg<sub>1</sub>, Seg<sub>2</sub>, ..., Seg<sub>L</sub>]

**Initialization:** Structured dictionary of labeled samples  $D (D_1, D_2, \dots, D_M)$ , sparsity level  $K_0$

**For** each pixel  $x$  in HSI data matrix set  $\mathbf{X}$

1. Find all superpixels  $\{sup_1\}, \{sup_2\}, \dots, \{sup_L\}$  containing  $x$  in all segmentations, the coverage regions of all superpixels are initialized as a complete matrix  $[Y^{sup-1}, \dots, Y^{sup-k}, \dots, Y^{sup-L}]$ ,  $k=1, 2, \dots, L$
2. Compute the correlation vector between test pixel  $x$  and  $Y^{sup-k}$ , then sort the pixel sequence as  $S^{sup-k} \{P_1, P_2, \dots, P_P\}$ , for each superpixels  $\{sup_1, sup_2, \dots, sup_L\}$  in descending by coefficient value
3. For all  $S^1, S^2, \dots, S^L$ , divide each  $S^{sup-k}$  into two parts as  $S_{com}^{sup-k}$  and  $S_{ind}^{sup-k}$ , and then select  $N$  pixels from  $S_{com}^{sup-k}$  and  $S_{ind}^{sup-k}$  as the reconstruction matrix  $Y^{SRM} \in \mathbb{R}^{P \times B}$  of test pixel  $x$  for combined region of superpixels
4. Compute the self-residual correlation matrix as  $R_x^t = D_C^T Y^{SRM}$
5. According to the correlation ranking in  $R_x^t$ , and select top  $N$  atoms merged the index  $I_{core}^{c,t}$  into index set  $I$
6. Sum the correlation degree of each class, to update the complete dictionary  $D_I$  for each class (1 to  $M$ )
7. Sparse coefficients are estimated according to the updated dictionary  $D_I$

$$A_{SRM}^t = (D_I^T D_I)^{-1} D_I^T Y^{SRM}$$

8. Compute the updated residuals correlation matrix

$$R_{t+1}^x = D^T (Y^{SRM} - D_I A_{SRM}^t)$$

9. While stopping criterion ( $t > K_0$ ) has not been met, repeat 4-8

10. Label the test pixel by

$$C_i = \arg \min_c \|Y_{SRM} - D_C A_{SRM}^t\|_F$$

11. Turn to the next test pixel

**End For**

**Output:** A 2-D matrix which records the labels of the all pixels

## 4. Experimental results

In this section, we evaluated the effectiveness of the proposed MSJSRC method on three real HSI datasets generated by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) and the Reflective Optics Spectrographic Imaging System (ROSIS) instruments. In addition, we compared the proposed method with some advanced classification methods.

### 4.1 Data Description and Experimental Setup

We selected three well-known publicly available data sets for experimental analysis, including Indian Pines, Salinas, and Pavia University. The class descriptions and sample distributions for the three data sets are given in [Table 1](#), [2](#) and [3](#). The false-color images of three test images are displayed in [Fig. 4](#).

Indian Pines is generated from Northern Indiana on June 12, 1992, that consists of  $145 \times 145$  pixels contained 16 land-cover classes with 220 contiguous spectral bands to which we have removed 20 noisy bands because of the water absorption bands. The 16 ground-truth classes of training and test are listed in [Table 1](#). About 10% of labeled data are used as training samples and the rest are used for testing are listed in [Table 1](#).

The second testing set is the Salinas Valley collected by the AVIRIS sensor over Salina Valley, California in 1998. This data set is in size of  $512 \times 217$  pixels with 224 contiguous spectral bands to which we have removed 20 noisy bands and finally 204 out of the 224 bands are used in our experiment. The 16 ground-truth classes are listed in [Table 2](#). About 1% of labeled data are used as training samples and the rest are used for testing are listed in [Table 2](#).

The Pavia University was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor during a flight campaign over Pavia University, northern Italy. The number of spectral bands is 103 and the image size is  $610 \times 340$  pixels. There are 9 classes and the original number of spectral bands is 115. 12 noisy bands are removed and finally 103 out of the 115 bands are used in our experiment. The 9 ground-truth classes and 100 labeled pixels for each class were randomly chosen for training and the rest are used for testing are listed in [Table 3](#).

**Table 1.** Training and test set for 16 classes in Indian Pines dataset

Class	Name	Training	test
1	Alfalfa	6	48
2	Corn-notill	144	1290
3	Corn-min	84	750
4	Corn	24	210

5	Grass/Pasture	50	447
6	Grass/Trees	75	672
7	Grass/Pasture-mowed	3	23
8	Hay-windrowed	49	440
9	Oats	2	18
10	Soybeans-notill	97	871
11	Soybeans-min	247	2221
12	Soybeans-clean	62	552
13	Wheat	22	190
14	Woods	130	1164
15	Building-Grass-Trees-Drives	38	342
16	Stone-steel Towers	10	85
	Total	1043	9323

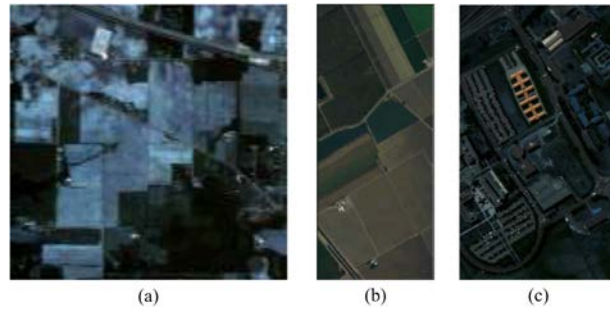
**Table 2.** Training and test set for 16 classes in Salinas dataset

Class	Name	Training	test
1	Brocoli_weeds	20	1989
2	Brocoli_weeds_2	37	3689
3	Fallow	20	1956
4	Fallow_rough_plow	14	1380
5	Fallow_smooth	27	2651
6	Stubble	40	3919
7	Celery	36	3543
8	Grapes_untrained	113	11158
9	Soil_vinyard_develop	62	6141
10	Corn_senesced_weeds	33	3245
11	Lettuce_romaine_4wk	11	1057
12	Lettuce_romaine_5wk	19	1908
13	Lettuce_romaine_6wk	9	907
14	Lettuce_romaine_7wk	11	1059
15	Vinyard_untrained	73	7195
16	Vinyard_trellis	18	1789
	Total	543	53586

**Table 3.** Training and test set for 9 classes in Pavia dataset

Class	Name	Training	test
1	Asphalt	100	6531
2	Meadows	100	18549
3	Gravel	100	1999
4	Trees	100	2964
5	Metal sheet	100	1245
6	Bare soil	100	4929
7	Bitumen	100	1230
8	Bricks	100	3582
9	Shadows	100	847
	Total	900	41876

To evaluate the performance of the proposed method, we adopted three commonly preferred objective metrics called overall accuracy (OA), average accuracy (AA), and the kappa coefficient(kappa)[37]. The overall accuracy shows the ratio between correctly classified test samples and the total number of test samples, and average accuracy shows the average of the percentages of the correctly classified samples in each class. The kappa coefficient is used to measure the correct degree of consistency of classified pixels. In addition, we also take into the consideration of the computation time and sparsity level during efficiency testing.

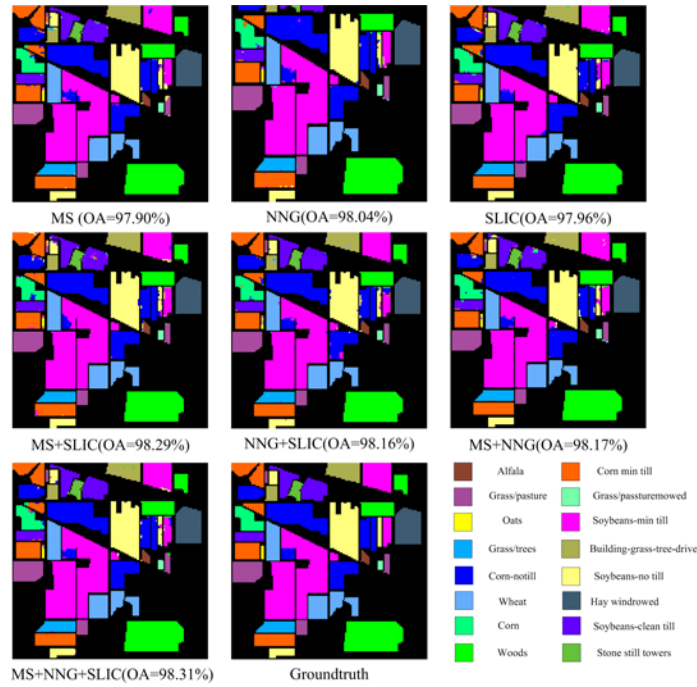


**Fig. 4.** False-color images of three test HSI: (a) AVIRIS Indian Pines image. (b)Salinas Valley data. (c) University of Pavia dataset.

## 4.2 Experiments on various superpixels

In this section, we test the proposed method on different superpixels sets obtained by three segmentation methods(Mean-shift, NNG, SLIC) and all their possible combinations as [MS, NNG, SLIC, MS+SLIC, MS+NNG, NNG+SLIC, MS+NNG+SLIC]. The experiment results of three test images are provided and analyzed followed.

We test the performance of the propose method on Indian Pines dataset at first, and 10% of the labeled data were randomly selected as the training samples. We conduct the experiment on three superpixels and the parameters setting is described in section 3.1. The number of discriminative optimal sampling is set as  $N=9$ . The classification maps obtained by various superpixels combination are shown in Fig. 5. From Fig. 5, we can see that the joint dictionary learning on multiple superpixels express more smooth area than single superpixels constraint. The quantitative metrics results (OA ,AA and Kappa coefficients) averaged over ten runs for various superpixels combinations are provided in Table 4. The overall classification accuracy (averaged over ten runs) is increased by 0.14% from [MS, NNG, SLIC] to [MS +SLIC, MS+NNG, NNG+SLIC], and overall classification accuracy (averaged over ten runs) is increased by 0.15\% from [MS +SLIC, MS+NNG, NNG+SLIC] to [MS+NNG+SLIC].



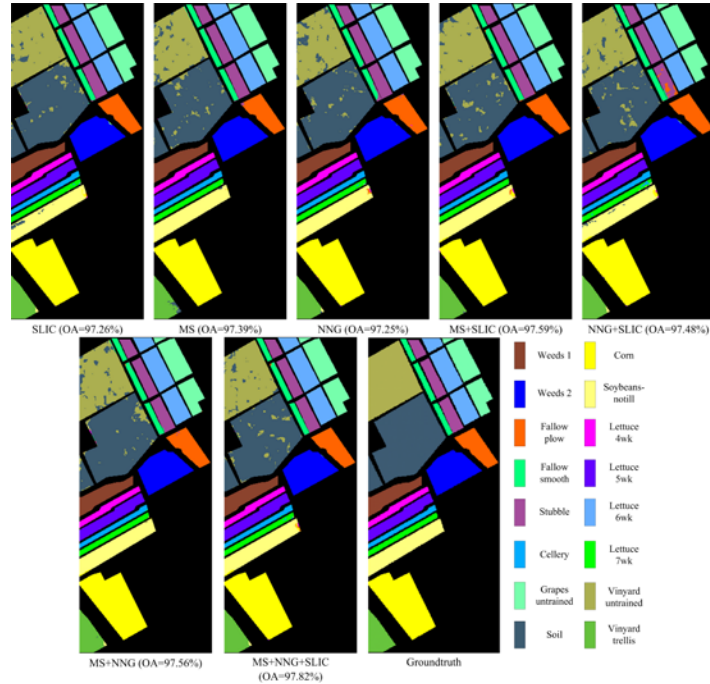
**Fig. 5.** Classification maps and overall classification accuracy (OA) for the Indian Pines dataset using about 10% labeled samples on various superpixels combinations.

**Table 4.** Classification accuracy (%) for the Indian Pines dataset on various superpixels combinations[MS, FH, SLIC, MS +SLIC, MS+ NNG, NNG+SLIC, MS+NNG+SLIC](N=9)

Class	MS	FH	SLIC	MS+SLIC	MS+NNG	NNG+SLIC	MS+NN G+SLIC
1	95.12	91.87	86.77	87.80	86.99	92.68	93.90
2	96.86	97.56	96.75	97.92	97.35	97.85	98.28
3	95.61	97.05	97.46	97.76	98.92	98.34	97.25
4	95.86	97.81	96.71	94.52	96.87	98.90	94.13
5	99.69	99.16	97.32	96.24	97.54	96.85	96.89
6	98.67	99.89	99.75	99.64	99.69	99.69	100.00
7	74.67	100.00	100.00	100.00	100.00	97.33	98.00
8	100.00	99.92	99.92	100.00	92.92	100.00	100.00
9	66.67	62.96	74.07	50.00	70.37	77.78	72.23
10	96.80	98.32	97.60	97.30	97.37	96.45	97.48
11	99.15	98.73	98.52	99.03	98.74	98.61	98.75
12	97.81	96.50	96.50	98.12	95.01	94.88	97.19
13	99.27	98.36	98.73	99.09	98.55	98.91	97.55
14	100.00	100.00	100.00	100.00	100.00	99.88	99.95
15	96.45	97.59	94.14	97.69	97.41	97.50	98.70
16	95.63	97.22	96.03	96.82	96.82	93.25	97.02
<b>OA</b>	98.12	98.17	97.87	98.23	98.22	98.13	98.34
<b>AA</b>	95.73	94.95	96.52	94.86	95.72	96.17	96.08
<b>Kappa</b>	97.14	97.99	97.61	97.98	97.97	97.87	98.10

For the Salinas image, 1% of the labeled data were randomly selected as the training samples. The number of discriminative optimal sampling is set as  $N=9$  in dictionary updating. The classification maps obtained by various superpixels combination are shown in Fig. 6. The quantitative metrics results (OA, AA and Kappa coefficients) averaged over ten runs for various superpixels combinations are provided in Table 5. The overall classification accuracy

(averaged over ten runs) is increased by 0.23% from [MS, NNF, SLIC] to [MS+SLIC, MS+NNG, NNG+SLIC], and overall classification accuracy (averaged over ten runs) is increased by 0.28% from [MS +SLIC, MS+NNG, NNG+SLIC] to [MS+NNG+SLIC]. We can see that the reconstruction area of the test pixel is changed with the joint constraints of superpixels, which expressed much more accuracy than single segmentation constraint.



**Fig. 6.** Classification maps and overall classification accuracy (OA) for the Salinas using about 1% labeled samples on various superpixels combinations.

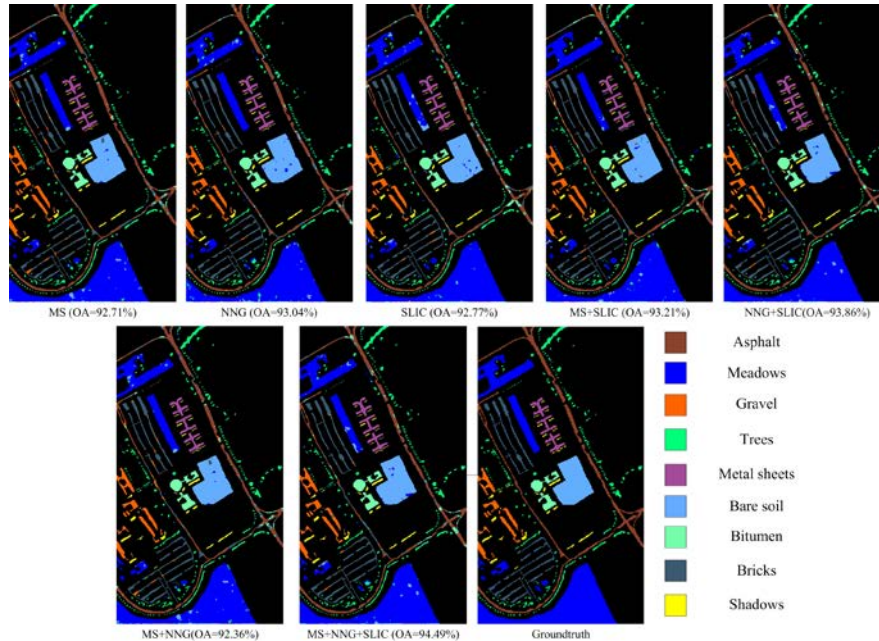
**Table 5.** Classification accuracy (%) for the Salinas dataset on various superpixels combinations [MS, FH, SLIC, MS +SLIC, MS+ NNG, NNG+SLIC, MS+NNG+SLIC] (N=9)

Class	MS	FH	SLIC	MS+SLIC	MS+NNG	NNG+SLIC	MS+NNG+SLIC
1	100.00	100.00	100.00	100.00	100.00	100.00	100.00
2	100.00	99.62	99.55	97.55	99.73	99.97	99.67
3	98.93	99.59	99.97	99.92	99.59	99.44	99.74
4	97.43	94.82	98.33	97.46	96.74	97.32	98.73
5	99.57	99.53	99.15	99.36	99.00	99.57	99.34
6	100.00	99.98	99.98	100.00	100.00	100.00	100.00
7	99.98	100.00	99.97	100.00	99.97	100.00	100.00
8	95.29	95.49	94.64	95.22	95.47	94.45	94.74
9	99.95	99.95	99.99	99.91	99.93	99.95	99.93
10	96.80	98.03	97.39	96.44	96.35	98.31	96.55
11	99.85	99.95	99.72	100.00	99.95	100.00	99.91
12	99.87	99.79	99.89	99.76	99.92	99.89	99.79
13	99.50	99.28	99.17	99.50	99.28	99.34	99.17
14	98.77	98.87	97.83	99.34	98.44	97.97	98.49
15	89.70	87.40	91.04	90.82	89.96	92.71	93.76
16	97.84	97.43	97.70	99.33	98.55	96.56	98.35
<b>OA</b>	97.19	96.92	97.26	97.38	97.22	97.47	97.64
<b>AA</b>	98.34	98.12	98.40	98.55	98.30	98.47	98.64
<b>Kappa</b>	96.87	97.05	96.95	97.08	97.05	97.19	97.38

For the Pavia University image, 100 labeled data were randomly selected as the training samples. The number of discriminative optimal sampling is set as  $N=12$ . The classification maps obtained by various superpixels combination are shown in Fig. 7. The quantitative metrics results(OA, AA and kappa coefficients) averaged over ten runs for various superpixels combinations are provided in Table 6. From Table 6, we can see that the overall classification accuracy (averaged over ten runs) is increased by 0.24% from [MS, NNF, SLIC] to [MS+SLIC, MS+NNG, NNG+SLIC], and overall classification accuracy (averaged over ten runs) is increased by 0.84% from [MS +SLIC, MS+NNG, NNG+SLIC] to [MS+NNG+SLIC]. The classification maps obtained by various superpixels combinations are shown in Fig. 7.

**Table 6.** Classification accuracy (%) for the Pavia University dataset on various superpixels combinations[MS, FH, SLIC, MS +SLIC, MS+ NNG, NNG+SLIC, MS+NNG+SLIC](N=9)

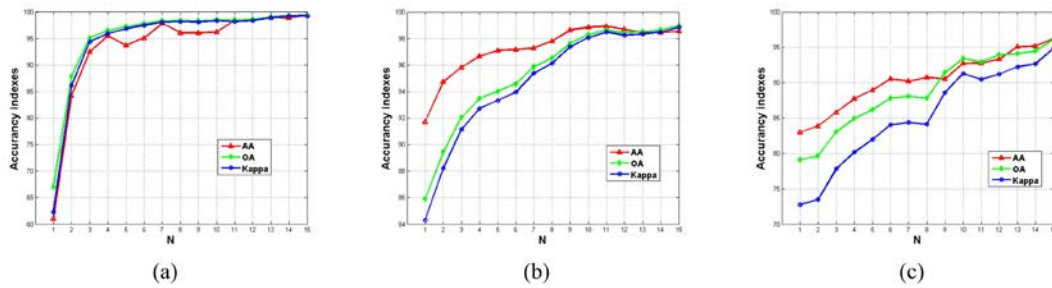
Class	MS	FH	SLIC	MS+SLIC	MS+NNG	NNG+SLIC	MS+NNG+SLIC
1	72.66	73.96	74.77	74.42	77.11	73.21	76.52
2	95.42	94.23	94.95	96.52	95.15	93.88	96.14
3	97.19	98.95	97.28	97.39	97.32	96.05	96.98
4	96.46	95.75	95.66	94.77	96.22	95.57	96.09
5	100.00	100.00	100.00	100.00	100.00	100.00	100.00
6	97.75	98.27	98.76	97.70	97.85	97.22	97.46
7	100.00	100.00	99.89	100.00	100.00	100.00	100.00
8	93.35	93.05	97.64	97.79	97.30	92.62	97.06
9	85.34	85.70	84.52	82.15	84.10	83.33	85.11
<b>OA</b>	92.19	91.46	92.73	93.17	93.08	91.32	93.36
<b>AA</b>	93.13	93.32	93.72	93.44	93.90	92.43	93.93
<b>Kappa</b>	89.70	89.39	90.41	90.96	90.86	88.56	91.22



**Fig. 7.** Classification maps and overall classification accuracy (OA) for the University of Pavia image using 100 labeled samples from each class on various superpixels combinations

The different scales of superpixels provide a relatively complete homogeneous region around test pixel, and coupled with discriminant optimal sampling can get complete and simplified reconstruction region. From the experiment results, we can see that more sampling pixels in reconstruction regions can take more characteristics of class and the classification is more accurate. Moreover, our method is convenient for the selection and adjustment of the reconstruction area which making the algorithm more efficient and feasible.

### 4.3 Effects of parameter N



**Fig. 8.** Classification effect of the selected atoms N for test images.(a) Indian pines; (b) Salinas; (c) University of Pavia

In this section, we further analyze the effects of parameter  $N$  on the three images. The selected segmental optimization atoms in dictionary updating are critical for joint representation, which is crucial for updating dictionary and generating adaptive sparse coefficients. The value of  $N$  has impacted on sparse representation, thus may affect the subsequent encoding process. We test the impact of  $N$  on the final classification accuracy of the MSJSRC on three test image. For the Indian Pines dataset, ten percent of the samples per class are chosen as the training set, and the remaining 90% are used for testing. For the Salinas dataset, one percent of the samples per class are chosen as the training set, and the remaining 99% are used for testing. For the Pavia University dataset, 100 labeled samples from each class are chosen as the training set, and the remaining are used for testing. For the parameter setting,  $N$  vary from 1 to 15 to illustrate the efficiency of classification accuracy. When  $N > 15$ , the calculation is quite time-consuming, and the accuracies on three test images are not improved obviously. How the OA, AA and Kappa coefficients acted as functions of  $N$  is shown in Fig. 8 respectively. We can see that variation tendencies are similar of the OA, AA and Kappa coefficients curve with different values of  $N$ . Concretely, as small  $N$  leads to low classification accuracy. As  $N$  increases, the sparse coding becomes more reliable and the classification accuracy increases.

### 4.4 Classification comparison with several excellent methods

To demonstrate the effectiveness of the proposed method, we compare it with those of six competing classification algorithms: PSRC[18], JSRM[18], MJSR[21], MASR[21], SC\_MK[22] and BTC[41]. In PSRC method, the sparse representation classification model is employed for classification with the spectral information only. The JSRM, MJSR and MASR method utilizes adopt the spatial context with fixed scales, multiple scales and adaptive multi-scales respectively. The selected region scales of JSRM is  $7 \times 7$ , and the various region scales selected as  $[3 \times 3, 5 \times 5, 7 \times 7, 9 \times 9, 11 \times 11, 13 \times 13]$  in MJSR and MASR, The sparsity level  $K=3$  for the MJSR and MASR methods as in original paper. The base superpixel number of SC\_MK is 800 as preset in[22]. BTC is a pixelwise classifier which uses only the spectral features and employs the WLS filters for smothing. The filtering process has two input



parameters: degree of smoothing ( $\lambda=0.4$ ) and the degree of sharpening ( $\alpha=0.9$ ) over the preserved edges.

The first test was performed on the Indian Pines image. We repeated 10 Monte Carlo runs to verify the average performance and the number of discriminative optimal sampling is set as  $N=10$ . The classification results of the seven algorithms are shown in Table 7. The classification maps with overall accuracies were provided in Fig. 8.

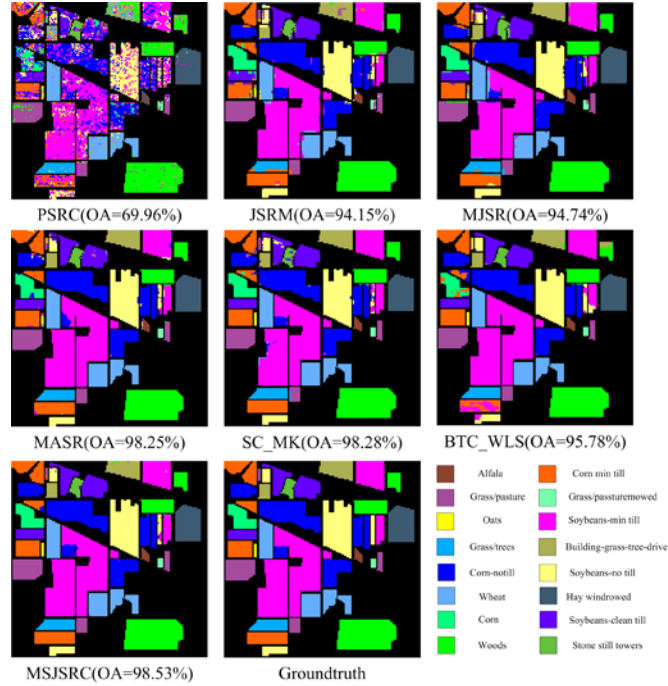


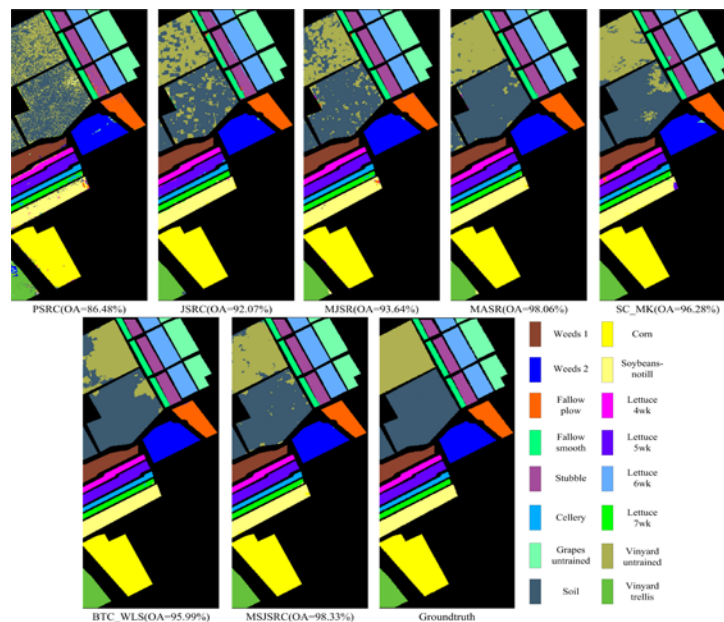
Fig. 8. Classification maps and overall classification accuracy (OA) of comparison methods on India image

Table 7. Classification accuracy(%) of comparison methods on India Pines images

Class	PSRC	JSRM	MJSR	MASR	SC MK	BTC-WSL	MSJSRC
1	39.84	84.55	83.74	94.31	100.00	99.07	94.31
2	54.29	92.42	93.64	98.19	97.43	93.02	97.20
3	52.21	92.41	93.44	97.50	98.06	89.91	98.35
4	41.31	92.64	88.41	96.09	97.87	90.92	94.99
5	85.36	94.33	94.41	96.24	97.35	94.32	97.70
6	89.75	93.71	97.92	99.95	100.00	100.00	99.95
7	74.67	85.33	84.00	100.00	100.00	98.15	88.00
8	96.28	99.85	99.77	99.85	100.00	100.00	100.00
9	24.08	31.48	59.26	66.67	100.00	100.00	79.63
10	69.37	91.70	93.22	95.47	93.48	92.26	97.60
11	70.88	96.73	97.13	98.98	99.14	99.17	99.25
12	39.76	92.57	88.23	94.38	97.56	99.19	98.179
13	91.85	80.98	93.84	98.55	99.73	100.00	98.55
14	90.48	98.54	99.56	100.00	100.00	98.92	100.00
15	41.11	91.64	89.15	97.60	97.69	95.48	98.75
16	88.10	88.49	86.90	96.43	89.28	99.20	96.83
<b>OA</b>	69.29	94.77	94.16	98.03	98.16	96.43	<b>98.53</b>
<b>AA</b>	65.58	88.21	91.05	95.64	97.98	96.85	<b>96.20</b>
<b>Kappa</b>	64.94	94.03	94.31	97.75	97.90	95.91	<b>98.33</b>

In **Fig. 8**, the PSRC displays a noisy class map by a single pixels sparse representation only using the spectral information. By combining the spatial information of the HSI, JSRM, MJSR, MASR, SC\_MK and BTC can exhibit significant smoother appearance in their classification results. The JSRM classifier exhibit a comparatively smooth result, and MJSR can improve the effect of JSRM with multi-scale local regions information. MASR can detect more meaningful complete regions than JSRM and MJSR. SC\_MK approach can not only provide a smoother appearance but also achieve more accurate estimations in the detailed area. When implementing the WLS filter, BTC approach can obtained better result in most areas. The proposed MSJSRC algorithm can also provide a better classification map with multiple superpixels constraint of spatial context. From the quantitative results in **Table 7**, the MSJSRC outperforms the comparative approaches in OA, AA and Kappa coefficients. We can observe that, MJSRC can obtained perfect accuracy in many areas except failure in some detailed areas.

In the experiments of the Salinas image, we also repeated 10 Monte Carlo runs to verify the average performance of the comparison methods, and The number of discriminative optimal sampling is set  $N=10$ . The classification results of the seven algorithms are shown in **Table 8**. The classification maps with overall accuracies are provided in **Fig. 10**. As can be seen in **Fig. 9**, the PSRC displays noisy class map only uses the spectral information. The performance of JSRM and MJSR are not increased remarkably in Vinyard\_untrained and Soil area by combining the spatial information with big scale region and multi-scales regions. MASR can deliver a smoother appearance with adaptive joint sparse than JSRM and MJSR in Vinyard\_untrained and Soil area. SC\_MK and BTC approach provide a smoother appearance than PSRC, JSRM and MJSR methods and achieve more accurate estimations in the Vinyard\_untrained area. Although these two approaches provide smoother appearance, but the classification estimations are not as accurate as MASR. The proposed MSJSRC algorithm deliver a better classification results than other compared methods in Vinyard\_untrained and Soil area. From the quantitative results in **Table 8**, the MSJSRC outperforms the comparative approaches in OA, AA and Kappa coefficients.



**Fig. 9.** Classification maps and overall classification accuracy (OA) of comparison methods on Salinas image

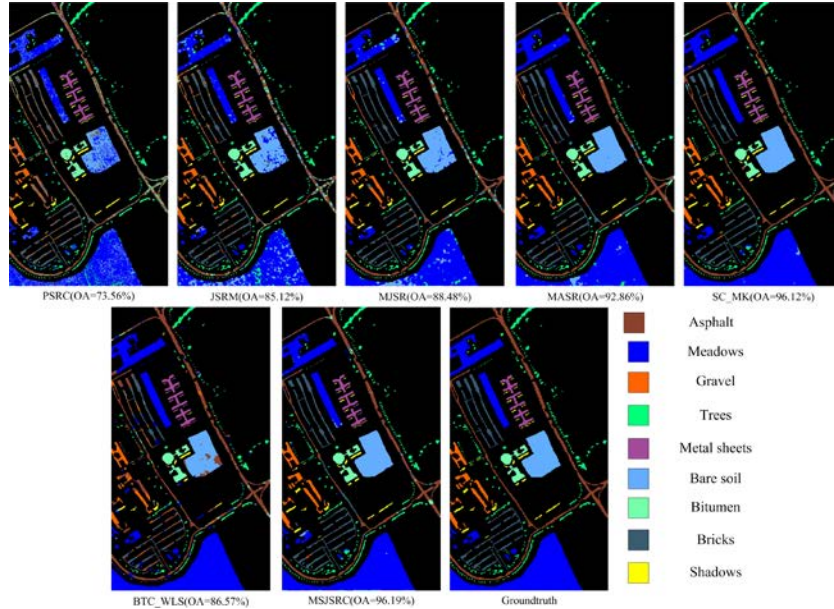
**Table 8.** Classification accuracy(%) of comparison methods on Salinas image

Class	PSRC	JSRM	MJSR	MASR	SC MK	BTC-WSL	MSJSRC
1	98.59	100.00	100.00	100.00	100.00	100.00	100.00
2	98.27	99.85	99.81	99.82	99.58	100.00	99.97
3	96.34	99.16	99.23	99.51	99.46	99.56	100.00
4	99.02	84.85	97.97	97.79	99.35	99.85	97.97
5	90.44	95.86	99.60	99.45	98.28	99.81	98.62
6	99.45	98.94	100.00	100.00	99.62	100.00	100.00
7	99.39	99.83	99.97	99.98	98.81	100.00	99.94
8	71.14	86.36	88.82	95.04	86.80	93.57	97.32
9	97.96	99.99	99.76	99.93	99.59	100.00	99.98
10	88.41	96.61	95.96	98.72	91.78	98.82	97.60
11	90.68	99.39	99.66	100.00	96.12	100.00	100.00
12	99.76	94.62	99.89	99.84	100.00	100.00	99.84
13	98.51	93.27	99.01	98.35	95.65	99.66	99.34
14	88.53	98.44	95.42	97.97	93.72	97.87	98.67
15	61.03	75.75	73.71	94.60	93.19	77.45	93.92
16	91.58	97.93	98.27	98.41	99.22	100.00	98.52
<b>OA</b>	86.22	92.56	93.60	97.91	95.86	95.48	<b>98.26</b>
<b>AA</b>	91.28	95.06	96.70	98.71	97.14	97.91	<b>98.85</b>
<b>Kappa</b>	84.66	91.71	92.86	97.69	95.39	94.97	<b>98.06</b>

In the experiments of the Pavia image, 10 Monte Carlo runs are repeated to verify the average performance of the comparison methods, and the number of discriminative optimal sampling is set  $N=15$ . The classification results of the seven algorithms are shown in **Table 9**. The classification maps with overall accuracies were provide in **Fig. 11**. We can see that the joint sparse model display improvements to a different extent than PSRC model, such as JSRM, MJSR, MASR and the proposed method. The JSRM classifier improves the visual effect than PSRC, MJSR and MASR can deliver of comparatively smooth results with multi-scale information of local regions. By combining the spatial information from superpixels, SC\_MK can display a smoother appearance in its classification results, while the BTC approach performs poorly on this data set. By contrast, the MSJSRC algorithm perform better than SC\_MK and MASR at meadows area. With multiple superpixels constraint of spatial context, it also provide a notable smooth classification map and achieve equivalent accuracy with SC\_MK approach. From the quantitative results in **Table 9**, the MSJSRC approach outperforms other compared approaches in OA, AA, and Kappa coefficients.

**Table 9.** Classification accuracy(%) of comparison methods on Pavia University dataset

Class	PSRC	JSRM	MJSR	MASR	SC MK	BTC-WSL	MSJSRC
1	60.10	60.22	60.21	75.92	95.59	60.21	87.30
2	78.43	85.78	89.18	95.88	95.10	99.49	97.56
3	90.65	93.93	95.14	96.10	98.27	63.53	96.76
4	90.65	93.93	95.14	96.10	98.27	63.53	96.76
5	99.36	99.96	99.96	100.00	99.88	100.00	99.96
6	66.58	86.20	94.57	98.92	97.01	86.33	99.19
7	87.27	98.98	100.00	100.00	96.64	99.25	100.00
8	70.19	87.95	95.45	95.84	97.32	53.33	97.87
9	98.10	74.53	76.54	84.10	100.00	42.90	86.86
<b>OA</b>	75.23	82.98	87.44	93.28	96.14	86.92	<b>96.16</b>
<b>AA</b>	80.12	86.23	90.04	93.95	97.29	78.20	<b>96.17</b>
<b>Kappa</b>	67.97	77.90	83.62	91.13	94.88	82.40	<b>94.89</b>

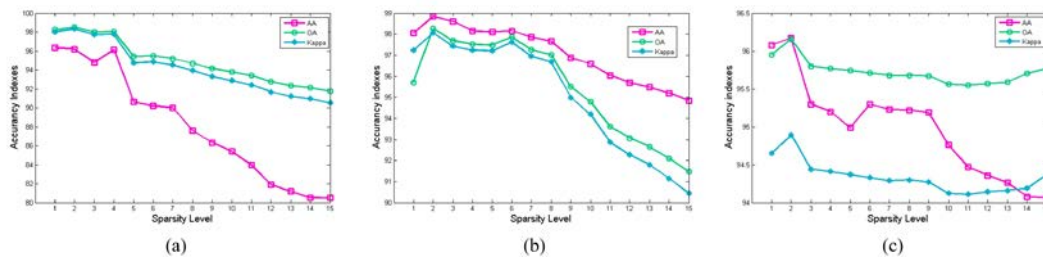


**Fig. 10.** Classification maps and overall classification accuracy (OA) of comparison methods on the University of Pavia image

In this section, we can observe that the MJSRC method can obtain perfect accuracy in many areas except failure in some detailed areas in India pines and Salinas datasets. The optimization sampling on multiscale superpixels acquire joint reconstructed region represent complete class attribute, and the joint sparse representation classification solving process can find the optimal sparse vector and dictionary sparsity. This method solves the blindness of spatial information fusion in sparse classification model and outperforms many classification algorithms.

#### 4.4 Effect of sparsity Level for the three test images

The effect of our method on various sparsity level ( $K$  vary from 1 to 15) is shown in Fig. 11. The OA, AA and Kappa coefficients show high performance at  $K=2$  of the proposed method under different sparsity levels for the three-test HSI images. By increasing the sparsity level from 3 to 15, the OA, AA and Kappa generally show a downward trend. So we select  $K=2$  in this paper for experiments as aforementioned.



**Fig. 11.** Effect of the sparsity level on three test images of our method. (a) Indian pines; (b) Salinas; (c) University of Pavia.

## 5. Conclusion

In this paper, we propose a HSI classification method based on the joint sparse representation with mixture spatial information of multi-layer superpixels. We extend the similar matrix of test pixel by optimally sampling with common and discriminative atoms from multi-layer superpixels. The advantages of this algorithm are that it can make full use of the spatial consistent information, and fully exploit the discriminative information of meaningful regions to develop the sparse coding. The experimental results conducted on publicly data sets show that our method has achieved better accuracy compared to several representative state-of-the-art methods. In the future, our work will apply kernel learning into joint sparse coding to optimize the classification performance.

## References

- [1] Zhang, E.; Zhang, X.; Jiao, L.; Li, L.; Hou, B. "Spectral-spatial hyperspectral image ensemble classification via joint sparse representation," *Pattern Recognition*, vol. 59, pp. 42-54, 2016. [Article \(CrossRef Link\)](#).
- [2] Bruzzone, L.; Chi, M.; and Marconcini, M. "A novel transductive SVM for semisupervised classification of remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, pp. 3363-3373, 2006. [Article \(CrossRef Link\)](#).
- [3] Zhong, Y.; Lin, Y.; and Zhang, L. "A support vector conditional random classifier with a Mahalanobis distance boundary constraint for high spatial resolution remotes sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, pp.1314-1333, 2014. [Article \(CrossRef Link\)](#).
- [4] Hao S, Wang W, Yan Y, et al. "Class-wise dictionary learning for hyperspectral image classification," *Neurocomputing*, vol. 220, pp. 121-129, 2017. [Article \(CrossRef Link\)](#).
- [5] Krishnapuram, B.; Carin, L.; Figueiredo, M.; and Hartemink, A. "Sparse multinomial logistic regression: Fast algorithms and generalization bounds," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 957-968, 2005. [Article \(CrossRef Link\)](#).
- [6] Li, J.; Bioucas-Dias, J. M.; and Plaza, A. "Semi-supervised hyperspectral image classification using soft sparse multinomial logistic regression," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, pp. 318- 322, 2013. [Article \(CrossRef Link\)](#).
- [7] Du B.; Zhang L.; Chen T.; Wu, K. "A discriminative manifold learning based dimension reduction method for hyperspectral classification," *International J Fuzzy Syst.*, vol. 14, pp.272-277, 2012. [Article \(CrossRef Link\)](#).
- [8] Li, W.; Prasad, S.; Fowler, J.E.; Bruce, L.M. "Locality-Preserving Dimensionality Reduction and Classification for Hyperspectral Image Analysis," *IEEE Transactions on Geoscience & Remote Sensing*, vol. 50, pp.1185-1198, 2012. [Article \(CrossRef Link\)](#).
- [9] Chen, Y.N.; Hsieh, C.T.; Wen, M. G.; Han, C.C.; Fan, K.C. "A Dimension Reduction Framework for HSI Classification Using Fuzzy and Kernel NFLE Transformation," *Remote Sensing*, vol. 7, pp. 14292-14326, 2015. [Article \(CrossRef Link\)](#).
- [10] Song, X.F.; Jiao, L.C.; Yang, S.Y.; Zhang, X. R.; Shang, F.H. "Sparse coding and classifier ensemble based multi-instance learning for image categorization," *Signal Process.* vol. 93, pp.1-11, 2013. [Article \(CrossRef Link\)](#).
- [11] Ham, J.; Chen; Y.C.; Crawford, M.M.; Ghosh, J. "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.* vol. 43, pp. 492-501, 2005. [Article \(CrossRef Link\)](#).
- [12] Wright, J., Yang, A.Y.; Ganesh, A.; Sastry, S.; and Ma, Y. "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, pp.210-227, 2009. [Article \(CrossRef Link\)](#).
- [13] Mairal, J.; Elad, M.; Sapiro, G. "Sparse representation for color image restoration," *IEEE Trans. Image Process.* vol. 17, pp. 53-69, 2008. [Article \(CrossRef Link\)](#).

- [14] Nasrabadi, N. M. "Hyperspectral Target Detection : An Overview of Current and Future Challenges," *IEEE Signal Process. Mag.*, vol. 31, pp. 34-44, 2013. [Article \(CrossRef Link\)](#).
- [15] Moayed, F.; Azimifar, Z.; Boostani, R. "Structured sparse representation for human action recognition," *Neuro computing*, vol. 161, pp. 38-46, 2015. [Article \(CrossRef Link\)](#).
- [16] Ma, W.K.; Bioucas-Dias, J.M.; Chan, T.H.; Gillis, N.; Gader, P.; Plaza, A.J.; Ambikapathi, A.; Chi, C.Y. "A signal processing perspective on hyperspectral unmixing," *IEEE Signal Process. Mag.* vol. 31, pp. 67-81, 2014. [Article \(CrossRef Link\)](#).
- [17] Chen, Y.; Nasrabadi, N.M.; Tran, T. D. "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Geosci. Remote Sens.* vol. 49, pp.3973-3985, 2011. [Article \(CrossRef Link\)](#).
- [18] Chen, Y.; Nasrabadi, N. M.; Tran T. D. "Hyperspectral image classification via kernel sparse representation," *IEEE Geosci. Remote Sens.*, vol. 51, pp. 217-231, 2013. [Article \(CrossRef Link\)](#).
- [19] Zhang, H.Y.; Li, J.Y.; Huang, Y.C.; Zhang, L.P. "A nonlocal weighted joint sparse representation classification method for hyper-spectral imagery," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.* vol. 7, pp.2056-2065, 2014. [Article \(CrossRef Link\)](#).
- [20] Peng J, Du Q. Robust Joint Sparse Representation Based on Maximum Correntropy Criterion for Hyperspectral Image Classification. *IEEE Transactions on Geoscience & Remote Sensing*, PP(99):1-13, 2017. [Article \(CrossRef Link\)](#).
- [21] Fang, L.Y.; Li, S.T.; Kang, X.D.; Benediktsson, J.A. "Spectral-spatial hyperspectral image classification via multiscale adaptive sparse representation," *IEEE Trans. Geosci. Remote Sens.* vol.52, pp.7738-7749, 2014. [Article \(CrossRef Link\)](#).
- [22] Fang, L.Y.; Li, S.; Duan, W.; Ren, J. "Classification of Hyperspectral Images by Exploiting Spectral Spatial Information of Superpixel via Multiple Kernels," *IEEE Trans. Geosci. Remote Sens.* 2015, vol. 53, pp.6663-6674. [Article \(CrossRef Link\)](#).
- [23] Fang L.Y., Li S.; Kang X.; Benediktsson, J.A. "Spectral-Spatial Classification of Hyperspectral Images With a Superpixel-Based Discriminative Sparse Model," *IEEE Trans. Geosci. Remote Sens.* vol. 53, pp.4186-4201, 2015. [Article \(CrossRef Link\)](#).
- [24] Peng, J.; Luo, T. "Sparse matrix transform-based linear discriminant analysis for hyperspectral image classification," *Signal, Image and Video Processing*, vol.10, pp.1-8, 2016. [Article \(CrossRef Link\)](#).
- [25] Zhai, Y.; Zhang, L.; Wang N.; Guo, Y. "A Modified Locality-Preserving Projection Approach for Hyperspectral Image Classification," *IEEE Trans. Geosci. Remote Sens. Lett.*, vol. 13, pp. 1-5, 2016. [Article \(CrossRef Link\)](#).
- [26] Li X, Liu K, Dong Y. "Superpixel-Based Foreground Extraction with Fast Adaptive Trimaps," *IEEE Transactions on Cybernetics*, PP(99):1-11, 2017. [Article \(CrossRef Link\)](#).
- [27] Sima H, Guo P. "Texture superpixels merging by color-texture histograms for color image segmentation," *Ksii Transactions on Internet & Information Systems*, 8(7):2400-2419, 2014. [Article \(CrossRef Link\)](#).
- [28] Dong Y, Tao D, Li X, et al. "Texture Classification and Retrieval Using Shearlets and Linear Regression," *IEEE Transactions on Cybernetics*, 45(3):358-369, 2015. [Article \(CrossRef Link\)](#).
- [29] Li X, Kang L, Dong Y, et al. Patch Alignment Manifold Matting. "*IEEE Trans Neural Netw Learn Syst.*" PP(99):1-13, 2017. [Article \(CrossRef Link\)](#).
- [30] Huang, H.; Yang, M. "Dimensionality Reduction of Hyperspectral Images With Sparse Discriminant Embedding," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, pp. 5160-5169, 2015. [Article \(CrossRef Link\)](#).
- [31] Felzenszwalb, P.F.; Huttenlocher, D. P. "Efficient Graph-Based Image Segmentation. *International Journal of Computer Vision*," *Int. J. comput. vis.*, vol. 59, pp. 167-181, 2004. [Article \(CrossRef Link\)](#).
- [32] Comaniciu, D.; Meer, P. "Mean shift: a robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.24, pp. 603-619, 2002. [Article \(CrossRef Link\)](#).
- [33] Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süsstru, S. "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.* vol. 34, pp. 2274-2282, 2012. [Article \(CrossRef Link\)](#).

- [34] Yuan, X.T.; Liu, X.; Yan, S. "Visual classification with multitask joint sparse representation". *IEEE Transaction on Image Process*, vol. 21, pp. 4349-4360, 2015. [Article \(CrossRef Link\)](#).
- [35] Duarte, M. F.; Sarvotham, S.; Baron, D.; Wakin, M.B. "Distributed Compressed Sensing of Jointly Sparse Signals," in *Proc. of Proceedings of the IEEE Conference of the Thirty-Ninth Asilomar Conference On signals, Systems and Computers*, pp. 1537-1541, 2005. [Article \(CrossRef Link\)](#).
- [36] Hu, Z.P.; Bai, F.; Zhao, S.H.; Wang, M.; Sun, Z. "Extended common molecular and discriminative atom dictionary based sparse representation for face recognition," *J of Vis. Commun. Image R.*, vol. 40, pp.42-50, 2016. [Article \(CrossRef Link\)](#).
- [37] Richards J. A., Jia X. P., Richards J A. "In Remote sensing digital image analysis: an introduction," *Remote sensing digital image analysis*, vol. 40, pp.47-54. 2016. [Article \(CrossRef Link\)](#).
- [38] Tropp, J.; Gilbert, A. "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, pp. 4655-4666, 2007. [Article \(CrossRef Link\)](#).
- [39] Chang S F, Wu X M, Li Z. "Segmentation using superpixels: A bipartite graph partitioning approach," in *Proc. of Proceedings of the IEEE Conference of the Computer Vision and Pattern Recognition*, June, pp. 789-796, 2012. [Article \(CrossRef Link\)](#).
- [40] Smith L. I. "A Tutorial on Principal Components Analysis," *Information Fusion*, vol. 51, 52, 2002. [Article \(CrossRef Link\)](#)
- [41] Töksoz, M.A.; Ilkay Ulusoy. "Hyperspectral Image Classification via Basic Thresholding Classifier," *IEEE Trans. on Geosci. Remote Sens.*, vol.54, pp. 4039-4051, 2016. [Article \(CrossRef Link\)](#).
- [42] Tropp J A. "Algorithms for simultaneous sparse approximation: part II: Convex relaxation[M]," *Elsevier North-Holland, Inc.* 2006. [Article \(CrossRef Link\)](#).



**Haifeng Sima** received the B.E. and M.E degree in computer science from Zhengzhou University, Zhengzhou, China, in 2004 and 2007 respectively, and the Ph.D. degree in software and theory from Beijing Institute of Technology, Beijing, in 2015. Since 2007, he has been with the faculty of Henan Polytechnic University, Jiaozuo, China, and is currently a lecturer with the School of Computer Science and Technology, Henan Polytechnic University. His current research interests include pattern recognition, image processing, image segmentation and image classification.



**Aizhong Mi** received the B.E. degree in computer application from HeFei University of Technology, Hefei, China, in 1999, and the Ph.D. degree in computer application technology from University of Science and Technology Beijing, Beijing, China, in 2009. Since 1999, he has been with the faculty of Henan Polytechnic University, Jiaozuo, China, and is currently an associate professor with the School of Computer Science and Technology, Henan Polytechnic University. His current research interests include pattern recognition and network security.



**Han Xue** received the B.E. degree in electronic commerce Henan Polytechnic University, Jiaozuo, China, in 2012 and 2015. She is a postgraduate student study at the School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, China. Her current research interests include pattern recognition, image processing, image segmentation and image classification.



**Shouheng Du** received the B.Sc. degree in Energy Science and engineering from Henan Polytechnic University, China in 1988. Currently, he is an associate professor at School of Computer Science and Technique, Henan Polytechnic University, China. His research interests include computer vision, pattern recognition, and image processing



**Zhiheng Wang** received the B.Sc. degree in mechatronic engineering from Beijing Institute of Technology, China in 2004, and the Ph.D. degree from Institute of Automation, Chinese Academy of Sciences, China in 2009. Currently, he is an associate professor at School of Computer Science and Technique, Henan Polytechnic University, China. His research interests include computer vision, pattern recognition, and image processing. E-mail:wzhenry@eyou.com



**Wang Jianfang** is an Associate Professor of the School of Computer Science and technology of the Henan Polytechnic University. He obtained PhD in School of Computer from University of the Northwestern Polytechnical University, in 2010. His research interests include recommender systems, deep learning, and social network analysis. The main research focus is to resolve the challenges of recommender systems, including matrix factorization, data sparsity, cold start, diversity and so on.