



# Hyperspectral image classification via compact-dictionary-based sparse representation

Chunhong Cao<sup>1</sup> · Liu Deng<sup>1</sup>  · Wei Duan<sup>1</sup> · Fen Xiao<sup>1</sup> · WanChun Yang<sup>1</sup> · Kai Hu<sup>1</sup>

Received: 27 February 2018 / Revised: 11 September 2018 / Accepted: 13 November 2018 /  
Published online: 22 November 2018  
© Springer Science+Business Media, LLC, part of Springer Nature 2018

## Abstract

In this paper, a compact-dictionary-based sparse representation (CDSR) method is proposed for hyperspectral image (HSI) classification. The proposed dictionary in CDSR is dynamically generated according to the spatial and spectral context of each pixel. It can effectively shrink the decision range for classification, and reduce the computational burden since the compact dictionary is composed of the classes correlated with the target pixel in terms of spatial location and spectral information. In order to obtain better spatial context information, a spatial location expanding strategy is designed for spreading local explicit label information to a wider region. Experimental results demonstrate the effectiveness and superiority of the proposed method when compared with some widely used HSI classification approaches.

**Keywords** Classification · Compact dictionary · Hyperspectral image · Sparse representation

## 1 Introduction

Compared with multispectral images, hyperspectral images contain more spectral information which consists of hundreds of different spectral bands, as shown in Fig. 1a. These continuous narrow spectral bands can reflect more detailed, potential, and discriminative land cover information [1, 3, 8]. As shown in Fig. 1b, it can present the spectral differences of different materials clearly on part of bands in HSI. In recent years, the HSI processing has been widely used in target detection [8], precision agriculture [13], military application [20], environmental monitoring [26] etc.

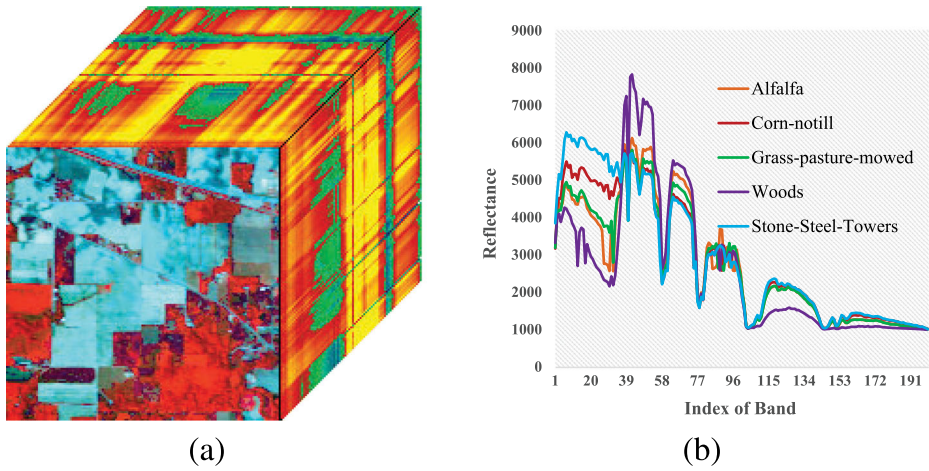
HSI classification plays a key role in HSI applications. It aims at assigning a specific label to each pixel so that it describes ground object more precisely [6, 14]. During the last

---

✉ Liu Deng  
dengliu-email@foxmail.com

Kai Hu  
kaihu@xtu.edu.cn

<sup>1</sup> Key Laboratory of Intelligent Computing and Information Processing of Ministry of Education, Xiangtan University, Xiangtan, China



**Fig. 1** (a) Visual hyperspectral data cube; (b) Spectral curve of some land covers in Indian Pines dataset

decade, many efforts have been made to complete this task. The SVM-based HSI classification frames are the most classical method, Camps-Valls and Melgani et al. make excellent contributions in this area [2, 23]. In addition, multinomial logistic regression [21], adaptive artificial immune network [19], gaussian process [31], the graph-based model [12, 29] and subject projection [38] are the commonly used classifier in HSI area. All these above mentioned approaches have achieved good classification performance in different aspects by combining their own advantages and the characteristics of HSI. Recently, deep learning architectures have been presented promising performance in various fields such as image segmentation [28], video analysis [37] and HSI classification. To the authors' knowledge, Chen et al. [5] first introduced the deep learning architecture into HSI classification by utilizing a multilayer stacked autoencoders to extract high-level features. Thereafter, many improved deep learning approaches were developed to obtain a more excellent performance. For instance, Zhang et al. [42] utilized recurrent neural network (RNN) to extract local spatial sequential features for HSI classification. Cheng et al. [7] further improved the classification accuracy by exploring the features generated from different layers in convolutional neural networks (CNN), and designing a unified metric learning based framework to alternately learn spectral-spatial features.

As we know, the superior of deep learning depends on large amount of labeled data. However, the reliable labeled samples are limited in HSI areas for the high cost of time and labor charge in labeling. As stated by Pan et al. [25], some traditional approaches still outperform the deep-learning based frame. As a traditional machine learning technique, sparse representation (SR) has made remarkable achievements in many research areas such as dimensionality reduction [15], image annotation [16], video semantic recognition [17]. The outstanding advantage of this technique is that it does not require large amount of samples to train the model, and no training is required in the residual-based classification model. Thus, many SR based methods are developed for HSI classification. For example, Chen et al. [4] introduced the Laplacian constraint and the joint representation of test sample to the sparsity model, which effectively utilized the neighboring spatial information. Yuan et al. [39] presented a SR method, which contained more discriminative information by utilizing the set-to-set distance. Fu et al. [11] proposed a shape-adaptive SR classification method,

which sufficiently exploited the spatial information through constructing a shape-adaptive local region for each target pixel. In general, the aforementioned SR based methods can be categorized as full-sized dictionary based SR. This type of dictionary is composed of the entire training samples and grouped by classes. Simultaneously, another type of techniques associated with dictionary learning were proposed [30, 33, 35], which learned a more representative dictionary from training samples.

Both the two types of SR methods have achieved good performances in terms of accuracy in HSI classification. However, when adopting the full-sized dictionary, the sparsity model may cost much time to solve the coefficients by greed pursuit algorithms, since each target pixel needs to compare with all atoms in the dictionary for searching the most approximate one within each iteration [4, 11, 36, 39]. For the dictionary learning methods, although the learned dictionary may have fewer atoms than full-sized dictionary, the learning procedure is time consuming. In addition, the learned dictionary is not grouped clearly by classes in general such that it cannot be directly used to make a decision for the classification, resulting in an extra classifier to deal with the obtained coefficients [30, 33].

Inspired by the issues mentioned above, we expect to design a simple and effective dictionary, which is both thinner than the full-sized dictionary and still grouped by classes. Moreover, we also pursue no complex updating and iterations in the processing. Accordingly, in this paper, we proposed a compact-dictionary-based SR method for HSI classification. Similar to the full-sized dictionary, the designed compact dictionary is still grouped by classes. While, it no longer consists of the entire training samples but some specific classes of training samples. These classes are determined by the known neighboring labels of target pixel and the spectral similarity between target pixel and the center of each class of training samples. Besides, a spatial location expanding strategy is developed to further exploit the spatial information. During the classification, the compact dictionary is adaptively determined for each test pixel and its size is less than the full-sized dictionary. Also, it is noted that no iteration is required in the formation of the compact dictionary when compared with dictionary learning methods. Therefore, the time consumption of CDSR for HSI classification would be reduced.

To sum up, the major contributions of our work are summarized as follows. 1) A customized compact dictionary for each test pixel is constructed to eliminate the interference of unrelated classes, and reduce the computational burden in the HSI classification. 2) A spatial location expanding classification strategy is designed: first, the unlabeled neighbors of labeled pixels are classified; then these unlabeled neighbors become labeled pixels; continue this process until all test pixels are classified. The expanding strategy can make full use of the spatial contextual information during the generation of compact dictionary.

The rest of the paper is organized as follows. In Section 2, some related works are briefly reviewed. In Section 3, the detailed descriptions of the proposed compact dictionary based sparse representation are provided. In Section 4, experimental results on three commonly used hyperspectral images are presented. Finally, a conclusion and future work are given in Section 5.

## 2 Related works

Most of HSI classification tasks completed by SR are on the basis of the observation that the hyperspectral pixels in the same class are located in a low-dimensional subspace in many cases [4, 41]. Consequently, each pixel  $x$  in a hyperspectral image can be sparsely represented by an appropriate dictionary constructed by or learned from training samples.

Let  $\mathbf{x} \in R^B$  be a  $B$ -dimensional spectral vector representing a pixel in HSI,  $\mathbf{D} \in R^{B \times N}$  be the dictionary and  $\boldsymbol{\alpha} \in R^N$  be the sparse coefficient vector, then the above process can be formulated as

$$\mathbf{x} = \mathbf{D}\boldsymbol{\alpha} \tag{1}$$

where  $B$  is the number of bands of HSI and  $N$  represents the number of atoms in  $\mathbf{D}$ .

In HSI, in view of the fact that the pixels within an appropriate size of window belong to the same class with high probability. Hence, using a region to replace the single spectral pixel would be better to describe the land cover information, and the (1) can be modified as

$$\mathbf{X} = \mathbf{D}\mathbf{A} \tag{2}$$

where  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_Q] \in R^{B \times Q}$  is neighboring matrix, which is stacked by neighboring spectral vectors within a  $\sqrt{Q} \times \sqrt{Q}$  window centered at target pixel  $x$ ;  $\mathbf{A} \in R^{N \times Q}$  is the corresponding coefficient matrix; for the dictionary, in [4], it is designed as  $\mathbf{D} = [\mathbf{D}_1, \dots, \mathbf{D}_c, \dots, \mathbf{D}_C] \in R^{B \times N}$ , where  $C$  represents the categories of land covers,  $N$  denotes the number atoms in dictionary,  $\mathbf{D}_c = [\mathbf{d}_1^c, \mathbf{d}_2^c, \dots, \mathbf{d}_{N_c}^c] \in R^{B \times N_c}$  is the subdictionary constructed by  $N_c$  training pixels of the  $c$ th class and  $\sum_{c=1}^C N_c = N$ . To obtain the sparse coefficient matrix  $\mathbf{A}$ , the (2) can be reformulated as the follow optimal problem:

$$\hat{\mathbf{A}} = \arg \min_{\mathbf{A}} \|\mathbf{X} - \mathbf{D}\mathbf{A}\|_F, \quad s.t. \quad \|\mathbf{A}\|_{row,0} \leq K \tag{3}$$

where  $\|\cdot\|_F$  denotes the Frobenius norm,  $K$  represents the sparse constraint, and  $\|\cdot\|_{row,0}$  is called joint sparse norm which guarantees that the nonzero rows have the same index. This means that the neighboring pixels and the target pixel share the same sparse support but have different weight value, because the pixels within the window are assumed as the same class. Through applying Simultaneous Orthogonal Matching Pursuit (SOMP) algorithm [4], an approximate solution can be achieved. Then, the class of target pixel  $x$  can be directly determined by the minimal total residuals:

$$Class(x) = \arg \min_c \|\mathbf{X} - \mathbf{D}_c \hat{\mathbf{A}}_c\|_F, \quad c = 1, 2, \dots, C. \tag{4}$$

The model reviewed above is the classical joint sparse representation classification (JSRC). After that, many efforts have been made, mainly focused on three aspects and corresponding to the three parts in (2), neighboring matrix  $\mathbf{X}$ , sparse coefficient matrix  $\mathbf{A}$ , and dictionary  $\mathbf{D}$ . The shape-adaptive joint sparse representation (SASR) [11], weighting prior sparse representation (WPSR) [32] and spatial-aware dictionary learning (SADL) [30], are the representative methods that follow the aforementioned three schemes respectively.

### 2.1 SASR

The SASR considers that the fixed-size square window used in JSRC may be inappropriate, since the center pixel and the neighboring pixels in the same square window may belong to different classes, especially in complex areas of HSI. Thus, SASR designed a shape-adaptive window  $\Omega$  for each pixel by utilizing the local polynomial approximation (LPA) filtering technique and the intersection of confidence intervals. Then, a shape-adaptive neighboring matrix can be constructed as  $\mathbf{X}_{sa} = [\mathbf{x}_1, \mathbf{x}_2, \dots], \mathbf{x}_i \in \Omega$ , and the SASR can be formulated as follows:

$$\hat{\mathbf{A}} = \arg \min_{\mathbf{A}} \|\mathbf{X}_{sa} - \mathbf{D}\mathbf{A}\|_F, \quad s.t. \quad \|\mathbf{A}\|_{row,0} \leq K. \tag{5}$$

The final class of test pixel can be determined by the minimal residual:

$$Class(x) = \arg \min_c \| \mathbf{X}_{sa} - \mathbf{D}_c \hat{\mathbf{A}}_c \|_F, \quad c = 1, 2, \dots, C. \tag{6}$$

### 2.2 WPSR

Similar to SASR, WPSR also noted the possible differences of pixels within the neighboring window, and it is not directly focused on the expression of neighboring matrix, but introducing weighting coefficient into sparse coefficient and relaxing the sparse constraint. The model of WPSR can be expressed as follows:

$$\min_{\mathbf{A}} \frac{1}{2} \| \mathbf{X} - \mathbf{D}\mathbf{A} \|_F^2 + \lambda_1 \| \mathbf{A} \|_1 + \lambda_2 \sum_{i,j} w_{i,j} \| \mathbf{a}_i - \mathbf{a}_j \|_2^2 \tag{7}$$

where  $\lambda_1$  and  $\lambda_2$  are the regularization parameters.  $w_{i,j}$  is the weight coefficient, used to measure the spectral similarity among pixels in neighboring window. The larger weight value indicates the higher similarity. In WPSR, the  $w_{i,j}$  is obtained by the sparse subspace clustering method [9]. Once the sparse coefficient matrix is solved, the label of test pixel can be determined by the similar minimal residual rule as (4) and (6).

### 2.3 SADL

Different from many SR based classifiers, whose dictionary is directly composed of the complete set of training samples and grouped by classes. SADL is a data-driven model, and its fundamental goal is building a representative and discriminative dictionary to better represent the test pixel. The dictionary of SADL is learned from the samples, rather than directly composed of them.

Let  $\{ \mathbf{y}_i \}_{i=1}^t$  denote the set of all spectral vector in HSI. To integrate the spatial information into the dictionary learning, the spectral vector is separated into some groups according the spatial context, i.e.,  $\mathbf{Y} = [\mathbf{Y}_1, \dots, \mathbf{Y}_g, \dots, \mathbf{Y}_G]$ , the details of SADL can be formulated as follows:

$$\min_{\mathbf{D}, \mathbf{A}} \| \mathbf{Y} - \mathbf{D}\mathbf{A} \|_F^2 + \sum_{g=1}^G \gamma_g \| \mathbf{A}_g \|_{2,1} \quad s.t. \quad \forall n \quad \| \mathbf{d}_n \|_2 \leq 1 \tag{8}$$

where  $\gamma_g$  is the regularization parameter for the  $g$ th group,  $\mathbf{A}_g$  is the corresponding sparse matrix for  $\mathbf{Y}_g$ , and  $\| \mathbf{A}_g \|_{2,1}$  denotes the sum of  $l_2$  norm of the rows  $\mathbf{A}_g$ . The dictionary  $\mathbf{D}$  and sparse matrix  $\mathbf{A}$  can be solved by an alternatively iterative scheme. While, it is noted that the label of test pixel cannot be determined by the minimal residual rule since the learned  $\mathbf{D}$  has no grouping structure. Accordingly, an extra classifier must be involved to deal with the sparse coefficient.

In this paper, our attention also focuses on the construction of the dictionary. In the construction procedure, the complex updating and iterator are avoided compared with the dictionary learning technique. And the constructed dictionary is expected to be more thinner than full-sized dictionary but still keep the grouped structure.

## 3 Proposed CDSR for HSI classification

In this section, we present the details of the proposed approach from three aspects, i.e., the determination of spatial label set, the determination of spectral set as well as the final classification model. The flowchart of the proposed CDSR is shown in Fig. 2.

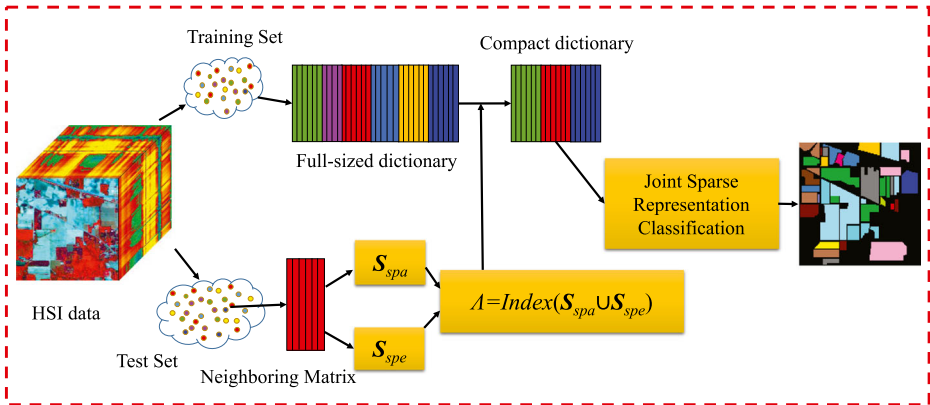


Fig. 2 The flowchart of the proposed CDSR

### 3.1 Determination of spatial label set

Generally, many spatial-spectral classifiers integrate the spatial information depending on the assumption that neighboring pixels tend to have similar spectral characteristics [10, 41]. However, they may ignore that the labels of neighboring pixels potentially provide the target pixel with an explicit reference. In this paper, the explicit information is used to shrink the size of the dictionary.

Through scanning a  $p \times p$  window centered at target pixel, it is easy to obtain a spatial label set  $S_{spa}$  which is composed of the known labels within the window. While limited by the number of training data and its spatial distribution, the obtained  $S_{spa}$  is an empty set sometimes if the test data is classified in random order. Therefore, a spatial location expanding strategy, which spreads the limited and local label information to a wider region, is proposed to deal with this case. The detailed expanding strategy is summarized to the following steps:

- Step 1: index the positions of pixels adjacent to the label-known pixels;
- Step 2: classify the unprocessed neighboring test pixels obtained by step 1;
- Step 3: repeat step 1 and 2 until all unprocessed test pixels are handled.

The graphic illustration of spatial location expanding strategy is shown in Fig. 3.

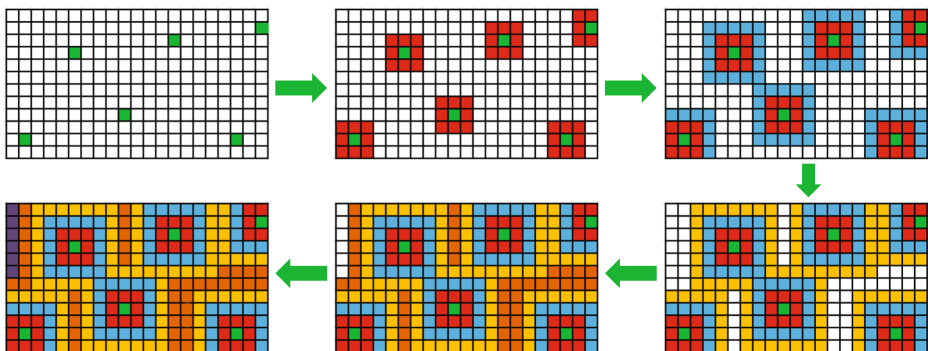


Fig. 3 Illustration of spatial location expanding strategy. Green blocks represent the training pixels, the blocks with other colors represent the test pixels that should be classified after each expansion

### 3.2 Determination of spectral label set

The obtained  $S_{spa}$  is beneficial to determine a relatively small scope for the compact dictionary, while it tends to be invalid in certain case. For instance, as shown in Fig. 4, the true label of target pixel is not contained in  $S_{spa}$  for there is no label-known pixels of the same class in the scanning window. And this case often appears in the regions that are not adjacent with training data. Under this case, the target pixels will suffer from misclassification. As a result, spectral information is introduced to obtain a more reliable scope for the compact dictionary.

Similar to  $S_{spa}$ , a spectral label set  $S_{spe}$  is obtained by measuring the spectral similarity between the neighboring matrix  $\mathbf{X}$  of the target pixel and each subdictionary  $\mathbf{D}_c$ . Since it is necessary for all test samples to measure the spectral similarity, a similar average strategy that has been also used in [34] is applied to  $\mathbf{X}$  and  $\mathbf{D}_c$  to speed up this procedure. Next, the detailed procedure of getting  $S_{spe}$  is given as follows:

Step 1: apply the average strategy to  $\mathbf{X}$  and  $\mathbf{D}_c$  to obtain the neighboring center  $\mathbf{u}$  and the subdictionary center  $\mathbf{v}^c, c = 1, 2, \dots, C$ , i.e.,

$$\mathbf{u} = \frac{1}{Q} \sum_{q=1}^Q \mathbf{x}_q \tag{9}$$

$$\mathbf{v}^c = \frac{1}{N_c} \sum_{i=1}^{N_c} \mathbf{d}_i^c, \quad c = 1, 2, \dots, C. \tag{10}$$

Step 2: use the spectral angle mapper (SAM) [43] to measure the spectral similarity between  $\mathbf{u}$  and  $\mathbf{v}^c$  for its simplicity and efficiency, i.e.,

$$sam(\mathbf{u}, \mathbf{v}^c) = \arccos \frac{\sum_{b=1}^B (u_b \times v_b^c)}{\sqrt{\sum_{b=1}^B (u_b)^2} \sqrt{\sum_{b=1}^B (v_b^c)^2}}, \quad c = 1, 2, \dots, C \tag{11}$$

where  $u_b^c$  and  $v_b^c$  denote the  $b$ th entry of  $\mathbf{u}$  and  $\mathbf{v}^c$ , respectively.

Step 3: sort all  $sam(\mathbf{u}, \mathbf{v}^c)$  with ascending order, and use the classes corresponding to the  $J$  smallest  $sam$  value to form the spectral label set  $S_{spe}$ , where  $J \leq C$  and  $J$  is called spectral similarity level for the convenience of following description.

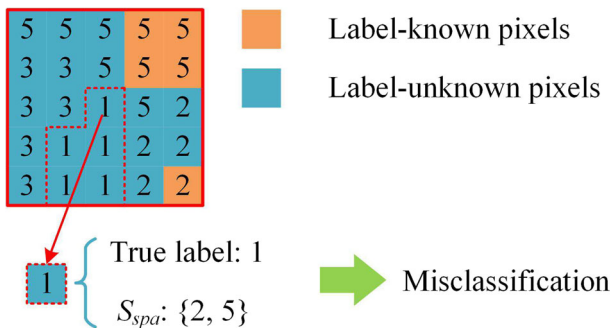


Fig. 4 Improper circumstances for determining the scope of compact dictionary just considering spatial information. (i.e.,  $S_{spa}$  does not contain the true label of test sample)



### 3.3 Incorporation of spatial and spectral information to design the compact dictionary

The final scope for target pixel is determined by the union of  $S_{spa}$  and  $S_{spe}$ ; and an index used to form the compact dictionary can be defined by

$$\Lambda = Index(S_{spa} \cup S_{spe}) \quad (12)$$

where  $Index(\cdot)$  is designed for getting the subdictionary index of corresponding class in full-sized dictionary  $\mathbf{D}$ . Then, the compact dictionary  $\tilde{\mathbf{D}}$  can be determined as

$$\tilde{\mathbf{D}} = \mathbf{D}_\Lambda \quad (13)$$

where  $\tilde{\mathbf{D}} \in R^{B \times \tilde{N}}$ ,  $\tilde{N}$  is the number of atoms in  $\tilde{\mathbf{D}}$ , and  $\mathbf{D}_\Lambda$  denotes the part of  $\mathbf{D}$  corresponding to the index  $\Lambda$ . Compared with the overcomplete dictionaries designed by many methods [4, 11, 39], the proposed compact dictionary turns to be undercomplete when the number of training data is small. In fact, as stated by [24], the overcomplete dictionary is not necessarily required for classification tasks, but plays a great role to reconstruction tasks. For example, methods proposed by [24] and [27] achieve the good classification performance when using the undercomplete dictionary.

After obtaining the compact dictionary, we use (3) and (4) with the corresponding compact dictionary  $\tilde{\mathbf{D}}$  to complete the classification task, i.e.,

$$\hat{\mathbf{A}} = \arg \min_{\mathbf{A}} \|\mathbf{X} - \tilde{\mathbf{D}}\mathbf{A}\|_F, \quad s.t. \quad \|\mathbf{A}\|_{row,0} \leq K \quad (14)$$

$$Class(x) = \arg \min_c \|\mathbf{X} - \tilde{\mathbf{D}}_c \hat{\mathbf{A}}_c\|_F, \quad c \in S_{spa} \cup S_{spe}. \quad (15)$$

For a more clear representation, the detailed procedure of CDSR is summarized as Algorithm 1.

---

#### Algorithm 1 CDSR

---

**Input:** 1) HSI; 2) coordinate of training samples; 3) neighboring window size  $\sqrt{Q}$  and scanning window size  $p$ ; 4) sparse constraint  $K$  and spectral similarity level  $J$ ;

**repeat**

1. Find the unprocessed test samples which are located around of label-known pixels, denoted as  $\mathbf{U}$ ;
- 2.

**for** each pixel  $x$  in  $\mathbf{U}$  **do**

- a. Obtain the spatial label set  $S_{spa}$  via the scanning window;
- b. Calculate the  $sam(\mathbf{u}, \mathbf{v}^c)$  via (11) and sort it with ascending order to form the spectral label set  $S_{spe}$ ;
- c. Determine the compact dictionary  $\tilde{\mathbf{D}}$  via (12) and (13);
- d. Apply  $\tilde{\mathbf{D}}$  to the joint sparsity model as follows:

$$\hat{\mathbf{A}} = \arg \min_{\mathbf{A}} \|\mathbf{X} - \tilde{\mathbf{D}}\mathbf{A}\|_F, \quad s.t. \quad \|\mathbf{A}\|_{row,0} \leq K$$

- e. Assign label for  $x$  based on the minimal residuals;

$$Class(x) = \arg \min_c \|\mathbf{X} - \tilde{\mathbf{D}}_c \hat{\mathbf{A}}_c\|_F, \quad c \in S_{spa} \cup S_{spe}.$$

**end for**

3. Remove these labeled test samples from test set;

**until** test set is empty

**Output:** a classification map of HSI.

---



## 4 Experiments and analysis

### 4.1 Comparison methods and metrics

To demonstrate the performance of the proposed CDSR, several related methods have been used for comparison, i.e. SVM [23], SVM-CK [2], OMP [4], JSRC [4], SASR [11], SADL [30] and NLSS-RNN [42]. SVM and SVM-CK are the classical and popular SVM-based HSI classifiers; OMP, JSRC, SASR and SADL are the SR-based methods, where SADL employs the technique of dictionary learning, while the others adopts the full-sized dictionary; NLSS-RNN is the popular deep-learning based approach, where a novel local spatial sequential feature was constructed and then was fed into the RNN for classification. In addition, the above methods can also be classified into two categories according to the use of spatial information. For a more clear representation, they are organized into a table as shown in Table 1. In this paper, three commonly used quantitative metrics are used to evaluate the performance of these methods. 1) Overall accuracy (OA): It denotes the ratio of correctly classified test samples in total test samples, i.e., the sum of correctly classified test samples divided by the total number of test samples. 2) Average accuracy (AA): It refers to the mean value of classification accuracy for all individual classes. 3) Kappa coefficient (KA): It is a statistic criteria in terms of omission and commission errors and can be used to measure the degree of agreement between classification result and ground truth. It can be computed by a confusion matrix, the detailed information is available in [18].

### 4.2 Experimental settings and results

For the SVM based methods, we choose the radial basis function (RBF) as the kernel function, the regularization term and kernel parameter are obtained by cross-validation technique. For OMP and JSRC, the optimal parameters are set to be the same as [4]. The parameters of SASR and SADL are also set as the same as the original paper [11, 30] respectively. For the proposed CDSR, the neighboring window size remains as the same as JSRC (i.e.,  $7 \times 7$ ,  $15 \times 15$ ,  $11 \times 11$  for Indian Pines, Salinas and Pavia University, respectively), since there is no difference in describing the test samples with neighboring information between JSRC and CDSR; the sparse constraint  $K$  is set to 3 by referring to [11]. The scanning window size and spectral similarity level  $J$  are determined by three-fold cross-validation. In the following, the optimal experimental results are presented by different methods in the case of above parameters on three HSI datasets, and all results are the average values on ten runs. For the NLSS-RNN, the detail data listed in following table are derived from the original paper [42].

**Indian Pines dataset** Indian Pines dataset was collected over the Indian Pines test site in North-western Indiana via Airborne/Visible Infrared Imaging Spectrometer (AVIRIS)

**Table 1** The characteristics of comparison methods

Types of methods	Spectral-spatial classifier	Only spectral classifier
SR-based (dictionary learning)	SADL	–
SR-based (full-sized dictionary)	JSRC, SASR	OMP
SVM-based	SVM-CK	SVM
Deep-learning-based	NLSS-RNN	–

sensor, the ground materials can be classified into 16 classes, the spatial size of this scene is  $145 \times 145$ , and its spatial resolution is 20 m. The original data contains 220 bands with the same size corresponding to the different narrow wavelength. 20 polluted bands were removed. Thus, only 200 bands were used in the experiment.

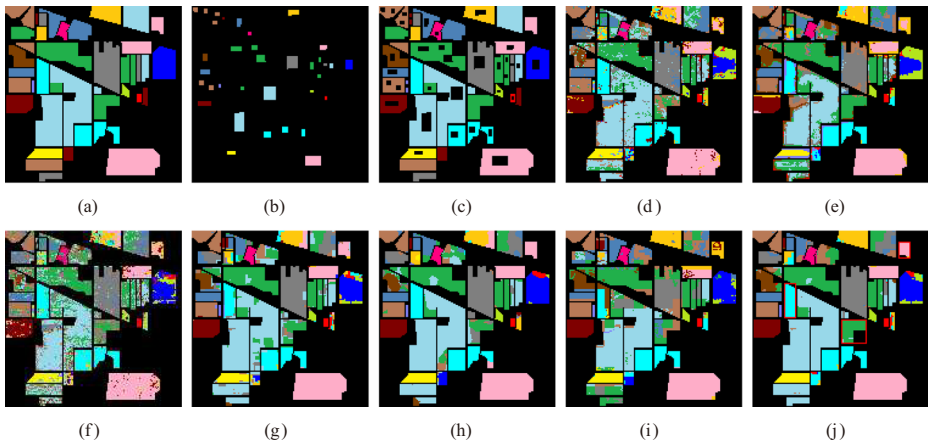
We randomly extract 10% labeled pixels from each class as the training set and the rest as the test samples, as shown in Table 2. The classification accuracies obtained by various classifiers are also shown in Table 2. As can be seen, the accuracies of spectral-spatial classifiers (i.e. SVM-CK, JSRC, SASR, SADL, NLSS-RNN) are obvious higher than that of the only spectral classifiers (i.e. SVM, OMP). The reason is that the integration of spatial information can enhance the robustness of single spectral pixel to noise. Moreover, it is easy to observe that the proposed CDSR outperforms the other classifiers in terms of OA, AA and KA. And in most land cover categories, CDSR achieves the highest accuracy among all these classifiers. Especially, there is no misclassification for class 1, 4, 6, 7, 8, 13 and 14. Such effectiveness of CDSR comes from the fact that the spatial location expanding strategy provides more precise scope for compact dictionary since the Indian Pines dataset is composed of many smooth regions. And the fewer classes contained in compact dictionary mean less interference from other unrelated classes to the sparse coefficients. Besides, the superior performance of CDSR can be partly attributed to that it not only exploits the label information of training samples, but also takes full advantage of the class information of the classified test samples during the classification.

In order to further verify the performance of the proposed CDSR, we test all mentioned methods using training samples obtained by manually block sampling as shown in Fig. 5b

**Table 2** Overall, average, and per class accuracy (%), and kappa coefficient of indian pines data set with different methods

Class	#Train	#Test	SVM	SVM-CK	OMP	JSRC	SASR	SADL	NLSS-RNN	CDSR
1	5	41	58.14	87.66	44.88	90.49	98.78	94.63	95.12	<b>100.00</b>
2	143	1285	73.99	90.73	55.81	95.42	97.03	93.91	98.21	<b>99.50</b>
3	83	747	72.14	92.15	51.06	93.20	98.53	95.15	<b>99.60</b>	98.15
4	24	213	62.67	85.23	37.56	93.47	93.99	91.97	97.18	<b>100.00</b>
5	48	435	85.41	94.99	82.92	93.98	98.00	96.44	97.93	<b>98.37</b>
6	73	657	89.48	96.77	91.58	95.92	97.91	99.19	98.33	<b>100.00</b>
7	3	25	77.90	79.45	73.60	85.20	98.80	<b>100.00</b>	96.00	<b>100.00</b>
8	48	430	94.85	98.41	93.98	99.05	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>
9	2	18	48.32	60.70	28.89	50.56	87.78	90.00	94.44	<b>97.78</b>
10	97	875	73.90	88.80	63.94	90.82	95.68	93.35	97.60	<b>99.98</b>
11	246	2209	77.67	92.02	71.38	96.30	98.74	97.85	99.28	<b>99.33</b>
12	59	534	77.96	90.98	42.77	89.01	95.97	91.63	98.31	<b>98.48</b>
13	20	185	90.63	98.66	90.97	88.43	99.14	99.41	<b>100.00</b>	<b>100.00</b>
14	126	1139	93.25	98.03	89.78	99.04	<b>100.00</b>	99.78	99.21	<b>100.00</b>
15	39	347	66.07	92.58	36.14	94.38	95.24	98.56	98.85	<b>99.86</b>
16	9	84	98.84	99.03	89.76	84.88	98.10	94.17	<b>100.00</b>	88.93
OA	–	–	79.98	92.90	68.77	94.77	97.89	96.53	98.75	<b>99.34</b>
AA	–	–	77.57	90.39	65.31	90.01	97.11	96.00	98.13	<b>98.77</b>
KA	–	–	0.758	0.917	0.655	0.933	0.976	0.968	0.990	<b>0.993</b>

The bold denotes the best result among all methods



**Fig. 5** Classification maps of Indian Pines in block sampling and the corresponding overall accuracy. (a) Ground Truth, (b) Training data, (c) Test data, (d) SVM (64.74%), (e) SVM-CK (62.38%), (f) OMP (59.02%), (g) JSRC (72.54%), (h) SASR (75.54%), (i) SADL (67.47%), (j) CDSR (79.32%)

(i.e., each class of training samples are gathered together and not distributed in all regions of the same class). It can be seen from Fig. 5d–j that many regions without training data suffer from misclassification for all methods. The main reason is that the selected training samples with the block sampling are unable to well cover the spectral variations of the same land-cover class in different regions and the mentioned spatial-spectral methods only utilize the local spatial information. However, CDSR still achieves the best results compared with other classifiers. Also, by introducing the spectral information, CDSR can handle those regions that are not adjacent with training data (e.g. most of pixels enclosed by red line are also recognized correctly as shown in Fig. 5j).

**Salinas dataset** The scene of Salinas dataset is located at Salinas Valley, California. 16 land covers were labeled in this area for groundtruth. The spatial size of the Salinas image is  $512 \times 217$ , and the spatial resolution of this image is 3.7 m per pixel. After discarding 20 water absorption bands, the remained 200 bands were used as the input of classifiers.

The Salinas dataset consists of many homogeneous regions and its spatial resolution is very high. In order to better demonstrate the superior performance of the proposed methods, we only extract 1% label-known pixels as training samples and the rest of label-known pixels are used to test the classification model as shown in Table 3. The visual classification maps are shown in Fig. 6. It can be observed that almost all classifiers make an obvious misclassification in the upper left area marked by yellow and blue of their corresponding classification maps except the proposed CDSR. The reason for this phenomenon lies in the fact that the spectral differences of these two area (corresponding to the untrained grapes and untrained vinyard in the land) are very small, and the local spatial information used in these spectral-spatial classifiers (i.e. SVM-CK, JSRC, SASR, SADL, CDSR) is beneficial to eliminate the influence of noise samples but it is not obvious for increasing the discrimination of spectral feature among land covers. While, for the CDSR, it can exclude many extraneous classes that are unrelated with target pixel in terms of spatial location from the decision range by using the designed compact dictionary, and the introduced spatial location expanding strategy can spread the local explicit label information to a wider region. Therefore, the CDSR obtains a more smoother classification map than other spectral-spatial

**Table 3** Overall, average, and per class accuracy (%), and kappa coefficient of salinas data set with different methods

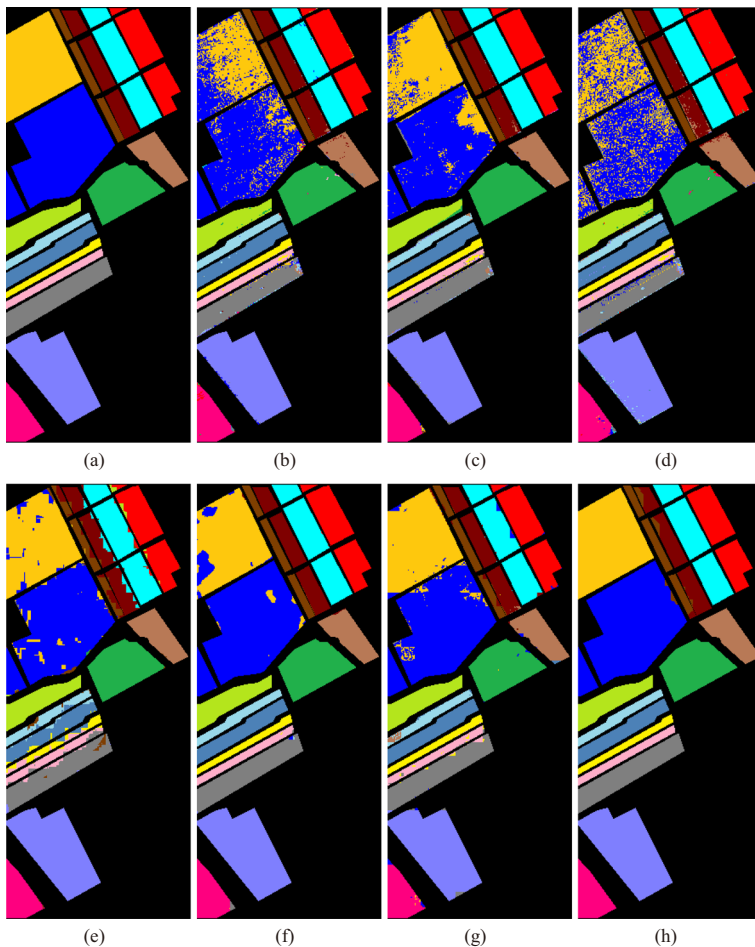
Class	#Train	#Test	SVM	SVM-CK	OMP	JSRC	SASR	SADL	NLSS-RNN	CDSR
1	20	1989	99.98	99.91	97.97	<b>100.0</b>	<b>100.0</b>	99.16	92.26	<b>100.0</b>
2	37	3689	98.23	99.44	98.50	99.90	99.82	99.63	99.76	<b>99.94</b>
3	20	1956	94.24	95.53	94.04	98.89	99.64	94.62	99.49	<b>99.97</b>
4	14	1380	97.46	97.56	99.21	70.30	<b>99.93</b>	95.35	99.78	97.45
5	27	2651	96.64	98.23	95.38	84.56	<b>98.88</b>	95.32	97.70	96.38
6	40	3919	99.46	99.97	99.49	94.29	99.97	99.46	99.77	<b>100.0</b>
7	36	3543	98.58	98.67	99.21	96.84	<b>99.99</b>	99.33	99.75	99.46
8	113	11158	78.81	89.89	72.50	96.48	97.52	95.99	94.76	<b>99.81</b>
9	62	6141	98.66	99.15	97.59	99.99	<b>100.0</b>	98.05	99.92	99.95
10	33	3245	90.74	91.11	85.66	94.10	<b>99.00</b>	92.87	96.95	97.90
11	11	1057	89.07	85.64	95.35	85.79	<b>100.0</b>	90.96	99.91	<b>100.0</b>
12	19	1908	97.61	97.75	99.37	86.54	99.98	96.87	99.90	<b>100.0</b>
13	9	907	95.75	96.78	97.57	74.49	99.66	91.25	98.79	<b>100.0</b>
14	11	1059	93.59	94.72	90.52	79.99	97.78	90.74	98.30	<b>100.0</b>
15	73	7195	76.57	84.84	58.89	93.20	94.39	95.56	93.01	<b>98.91</b>
16	18	1789	98.83	98.43	92.97	<b>99.63</b>	99.14	96.43	99.89	99.54
OA	–	–	90.53	94.14	86.30	94.19	98.51	96.49	97.23	<b>99.38</b>
AA	–	–	94.01	95.48	92.14	90.94	99.11	95.72	98.12	<b>99.33</b>
KA	–	–	0.894	0.935	0.847	0.935	0.983	0.961	0.970	<b>0.993</b>

The bold denotes the best result among all methods

approaches. Moreover, the detailed data for quantitative metrics is presented in Table 3. It is obvious that the proposed method outperforms other algorithms. Even compared to SASR, SADL and NLSS-RNN, the overall accuracy of CDSR is increased by 0.87%, 2.89% and 2.15%, respectively.

**University of Pavia dataset** Pavia University (PaviaU) dataset, which captured by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor, is of size  $610 \times 340$  with 115 spectral bands, 12 noisy bands among all bands were removed so that we can alleviate its negative effect to classification. This dataset possesses very high spatial resolution (1.3-meter pixels), and a groundtruth labeled with 9 classes is provided.

By referring to [42], we randomly extract 9% labeled pixels from each class as a balanced training set, and the rest of 91% as the test set (see Table 4). The ground truth map and classification maps of University of Pavia are displayed in Fig. 7. It can be seen that the Pavia University dataset consists of many long and scattered regions which are more complicated than the previous two datasets. Looking closely at Fig. 7b and d, there are many salt and pepper noise points in the homogenous regions, because the corresponding classifiers only take the spectral information into consideration. Furthermore, we can observe that SASR, SASL and CDSR obtain slightly difference classification maps. All of them are very close to the ground truth map. Hence, for a more effective comparison, the detailed quantitative metrics are also given in Table 4. It is easy to find that CDSR achieves the highest



**Fig. 6** Groundtruth map and classification maps of different methods on Salinas dataset (1% training data for each class in random sampling). (a) Ground Truth, (b) SVM, (c) SVM-CK, (d) OMP, (e) JSRC, (f) SASR, (g) SADL, (h) CDSR

accuracy in class 1, 2, 6 and 8. Also, it should be noted that CDSR is not as well as NLSS-RNN in terms of OA, AA and KA. The reason may be that the small and scattered regions in PaviaU are conducive to the spatial expanding strategy, and a relative large neighboring window size was set in CDSR for balancing other regions. While, the neighboring window size of this setting is not suitable for the small and scattered area. Due to this reason, class 4 and 9 get a relative lower results. In spite of this, compared with the SR-based methods, the CDSR still achieves a comparative result for this dataset in overall.

### 4.3 Computational complexity and running time

For the proposed CDSR, the time-consuming steps consist of three parts: searching scan window, calculating spectral angle mapper, and solving sparse coefficient by SOMP algorithm. For a classification task with  $M$  test samples, the computation complexity of determining spatial label set by searching scan window is  $O(p^2M)$ . The time complexity

**Table 4** Overall, average, and per class accuracy (%), and kappa coefficient of university of pavia data set with different methods

Class	#Train	#Test	SVM	SVM-CK	OMP	JSRC	SASR	SADL	NLSS-RNN	CDSR
1	597	6034	94.44	97.47	76.04	98.21	99.20	98.59	99.67	<b>100.00</b>
2	1678	16971	96.11	99.47	95.13	99.97	99.99	99.79	99.90	<b>100.00</b>
3	189	1910	85.11	94.42	62.12	98.78	99.65	99.25	<b>99.69</b>	99.09
4	276	2788	96.81	99.63	84.35	93.26	87.18	97.65	<b>99.21</b>	91.27
5	121	1224	99.57	99.46	99.61	98.28	<b>100.00</b>	99.74	<b>100.00</b>	97.22
6	453	4576	92.79	98.98	55.79	99.89	99.96	99.80	99.85	<b>99.97</b>
7	120	1210	87.55	96.91	79.17	98.53	<b>100.00</b>	99.40	99.34	98.47
8	331	3351	85.76	92.18	77.22	98.40	99.47	98.37	99.67	<b>99.99</b>
9	85	862	99.97	<b>100.00</b>	95.29	71.83	96.44	99.98	<b>100.00</b>	92.98
OA	–	–	94.03	98.17	83.26	98.29	98.81	99.29	<b>99.77</b>	99.04
AA	–	–	93.12	97.61	80.52	95.24	97.99	99.17	<b>99.70</b>	97.67
KA	–	–	0.922	0.974	0.772	0.979	0.987	0.991	<b>0.997</b>	0.989

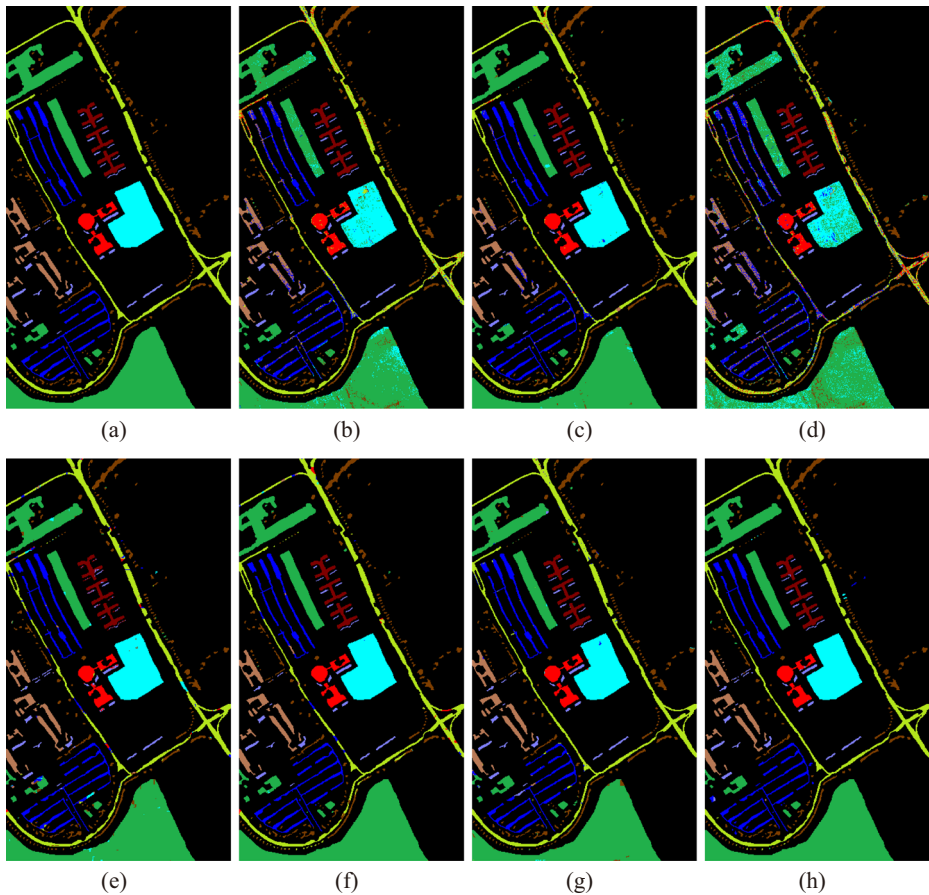
The bold denotes the best result among all methods

of calculating spectral angle mapper is  $O(3CMB)$ . By referring to [11], we know that the main operation of SOMP lies in the scalar multiplication, and its number of computation is  $\sum_k (QB(\tilde{N} - k) + 2k^2B + k^3 + kBQ)$ ,  $k = 1, 2, \dots, K$ . Therefore, the final computation complexity is  $O(M(QKB\tilde{N} + 2K^3B + K^2BQ + 3CB + p^2))$ .

In this part, we also compare the running time of the mentioned methods in 4.1 with the environment of MATLAB R2014a, Intel Core i5-4590 CPU 3.30GHz and 4GB RAM. Because the NLSS-RNN needs another totally different environment and the original paper does not provide the corresponding data, the running time of NLSS-RNN is not listed in Table 5 along with other methods. It can be seen that the SVM-based methods is less time-consuming, and the C interface called in its source code make a certain contribution. In addition, it should also be noted that the proposed CDSR is the fastest among all SR-based spectral-spatial classifiers including the full-sized dictionary based JSRC, SASR as well as the dictionary learning based SADL. The reason is that the compact dictionary greatly shortens the solution time of the sparse coefficient, and there are no additional iteration and update operation required in the construction of compact dictionary. Although, there is no detailed data about the time-consuming of NLSS-RNN, while, we know that the NLSS-RNN requires 1000 iterations to ensure the convergence of the algorithm according to the original paper [42]. Therefore, the training time of NLSS-RNN should not be too short. Overall, the proposed CDSR is efficient.

#### 4.4 Influence of parameters

Both scanning window size and spectral similarity level are the key parameters to influence the scale of compact dictionary. Thus, we analyze the influence of the two parameters in terms of classification accuracies on Indian Pines, Salinas and Pavia University datasets. From Fig. 8a, it can be seen that with the increasing size of the scanning window, the overall accuracy decreases generally; in particular, when the scanning window size is expanded to  $7 \times 7$  or more larger, the accuracy varies little. This can be explained that the compact dictionary contains bigger number of classes with the increasing size of the scanning window, and when the window size is large enough, the number of classes contained in the



**Fig. 7** Groundtruth map and classification maps of different methods on University of Pavia dataset (9% training data for each class in random sampling) (a) Ground Truth, (b) SVM, (c) SVM-CK, (d) OMP, (e) JSRC, (f) SASR, (g) SADL, (h) CDSR

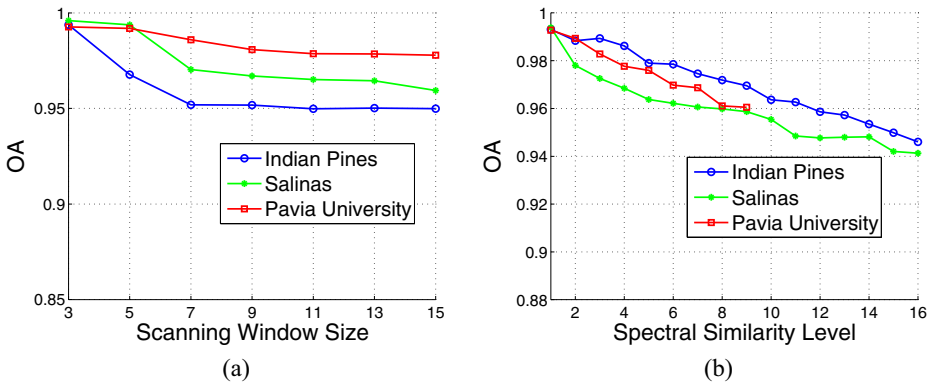
dictionary will not fluctuate greatly. From Fig. 8b, it can be observed that with the higher of spectral similarity level, the classification accuracy is gradually decreased, which can be explained by the fact that the compact dictionary degrades to be full-sized dictionary with the increasing spectral similarity level.

**Table 5** The average running time (seconds) for ten runs on indian pines, salinas and university of pavia datasets

Dataset	SVM	SVM-CK	OMP	JSRC	SASR	SADL	CDSR
IndianPines	<b>1.08</b>	2.52	6.63	31.62	34.03	106.43	12.19
Salinas	<b>3.42</b>	10.63	29.50	385.12	225.56	188.39	141.64
PaviaU	<b>8.37</b>	16.80	32.96	445.42	679.67	266.82	135.34

The bold denotes the best result among all methods





**Fig. 8** Overall accuracy with varying scanning window size and spectral similarity level on Indian Pines, Salinas and Pavia University

## 5 Conclusion

In this paper, we proposed a HSI classification method based on compact-dictionary-based sparse representation by uniting the spatial location expanding strategy and the spectral information. Experiments on three HSI datasets verify the performance of the proposed method in terms of accuracy and efficiency. In the future work, we will try to further improve the classification accuracy in block sampling by combining the global spatial information to eliminate the influence caused by the spectral variations of the same land-cover class in different regions.

**Acknowledgements** This work is supported by the National Natural Science Foundation of China under Grant No.61401386 and 61802328. The authors would like to thank Prof. D. Landgrebe and Prof. P. Gamba for providing the hyperspectral data set. In addition, we also would like to thank the reviewers for their constructive comments on this manuscript.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

1. Bioucas-Dias JM, Plaza A, Camps-Valls G, Scheunders P (2013) Hyperspectral remote sensing data analysis and future challenges. *IEEE Geosci Remote Sens Mag* 1(2):6–36
2. Camps-Valls G, Gomez-Chova L, Munoz-Mari J, Vila-Frances J, Calpe-Maravilla J (2006) Composite kernels for hyperspectral image classification. *IEEE Geosci Remote Sens Lett* 3:93–97
3. Camps-Valls G, Tuia D, Bruzzone L, Benediktsson JA (2014) Advances in hyperspectral image classification: earth monitoring with statistical learning methods. *IEEE Signal Proc Mag* 31(1):45–54
4. Chen Y, Nasrabadi NM, Tran TD (2011) Hyperspectral image classification using dictionary-based sparse representation. *IEEE Trans Geosci Remote Sens* 49(10):3973–3985
5. Chen Y, Lin Z, Zhao X, Wang G, Gu Y (2014) Deep learning-based classification of hyperspectral data. *IEEE J Selected Topics Appl Earth Observ Remote Sens* 7(6):2094–2107
6. Chen C, Chen N, Peng J (2016) Nearest regularized joint sparse representation for hyperspectral image classification. *IEEE Geosci Remote Sens Lett* 13(3):424–428
7. Cheng G, Li Z, Han J, Yao X, Guo L (2018) Exploring hierarchical convolutional features for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* pp(99):1–11

8. Du Q, Zhang L, Zhang B, Tong X, Du P, Chanussot J (2013) Foreword to the special issue on hyperspectral remote sensing: theory, methods, and applications. *IEEE J Selected Topics Appl Earth Observ Remote Sens* 6(2):459–465
9. Elhamifar E, Vidal R (2009) Sparse subspace clustering. *IEEE Comput Vis Pattern Recognit* 35:2790–2797
10. Fang L, Li S, Kang X, Benediktsson JA (2014) Spectral-spatial hyperspectral image classification via multiscale adaptive sparse representation. *IEEE Trans Geosci Remote Sens* 52(12):7738–7749
11. Fu W, Li S, Fang L, Kang X, Benediktsson JA (2016) Hyperspectral image classification via shape-adaptive joint sparse representation. *IEEE J Selected Topics Appl Earth Observ Remote Sens* 9(2):556–567
12. Gao Y, Ji R, Cui P, Dai Q (2014) Hyperspectral image classification through bilayer graph-based learning. *IEEE Trans Image Process* 23(7):2769–2778
13. Gevaert CM, Suomalainen J, Tang J, Kooistra L (2015) Generation of spectral temporal response surfaces by combining multispectral satellite and hyperspectral uav imagery for precision agriculture applications. *IEEE J Selected Topics Appl Earth Observ Remote Sens* 8(6):3140–3146
14. Ghamisi P, Plaza J, Chen Y, Li J, Plaza A (2017) Advanced spectral classifiers for hyperspectral images: a review. *IEEE Geosci Remote Sens Mag* 5(1):8–32
15. Han Y, Wu F, Tao D, Shao J, Zhuang Y, Jiang J (2012) Sparse unsupervised dimensionality reduction for multiple view data. *IEEE Trans Circuits Syst Video Technol* 22(10):1485–1496
16. Han Y, Wu F, Tian Q, Zhuang Y (2012) Image annotation by input-output structural grouping sparsity. *IEEE Trans Image Process* 21(6):3066–3079
17. Han Y, Yang Y, Yan Y, Ma Z, Sebe N, Zhou X (2015) Semisupervised feature selection via spline regression for video semantic recognition. *IEEE Trans Neural Netw Learn Syst* 26(2):252–264
18. Jia S, Deng B, Zhu J, Jia X, Li Q (2017) Superpixel-based multitask learning framework for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 55(5):2575–2588
19. Jiao H, Zhong Y, Zhang L (2012) Artificial DNA computing-based spectral encoding and matching algorithm for hyperspectral remote sensing data. *IEEE Trans Geosci Remote Sens* 50(10):4085–4104
20. Lampropoulos GA, Liu T, Qian SE, Fei C (2009) Hyperspectral classification fusion for classifying different military targets. *IEEE Int Geosci Remote Sens Symp* 3:262–265
21. Li J, Bioucas-Dias JM, Plaza A (2012) Spectralspatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields. *IEEE Trans Geosci Remote Sens* 50(3):809–823
22. Liu J, Xiao Z, Xiao L (2016) Superpixel-guided multiscale kernel collaborative representation for hyperspectral image classification. *Remote Sens Lett* 7(10):975–984
23. Melgani F, Bruzzone L (2004) Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans Geosci Remote Sens* 42(8):1778–1790
24. Moody DI, Smith DA, Brumby SP (2014) Automatic detection of pulsed radio frequency (rf) targets using sparse representations in undercomplete learned dictionaries. *Proc SPIE Automatic Target Recognition XXIV* 9090:C1–C11
25. Pan B, Shi Z, Xu X (2017) R-VCANet: a new deep-learning-based hyperspectral image classification method. *IEEE J Selected Topics Appl Earth Observ Remote Sens* 10(5):1975–1986
26. Pontius J, Martin M, Plourde L, Hallett R (2008) Ash decline assessment in emerald ash borer-infested regions: a test of tree-level, hyperspectral technologies. *Remote Sens Environ* 112(5):2665–2676
27. Roscher R, Waske B (2016) Shapelet-based sparse representation for landcover classification of hyperspectral images. *IEEE Trans Geosci Remote Sens* 54(3):1623–1634
28. Saha M, Chakraborty C (2018) Her2net: a deep framework for semantic segmentation and classification of cell membranes and nuclei in breast cancer evaluation. *IEEE Trans Image Process* 27(5):2189–2200
29. Shao Y, Gao C, Sang N (2017) A discriminant sparse representation graph-based semi-supervised learning for hyperspectral image classification. *Multimed Tools Appl* 76(8):1–13
30. Soltani-Farani A, Rabiee HR, Hosseini SA (2015) Spatial-aware dictionary learning for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 53(1):527–541
31. Sun S, Zhong P, Xiao H, Wang R (2014) Active learning with Gaussian process classifier for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 53(4):1746–1760
32. Sun X, Qu Q, Nasrabadi NM, Tran TD (2014) Structured priors for sparse-representation-based hyperspectral image classification. *IEEE Geosci Remote Sens Lett* 11(7):1235–1239
33. Sun X, Nasrabadi NM, Tran TD (2015) Task-driven dictionary learning for hyperspectral image classification with structured sparsity constraints. *IEEE Trans Geosci Remote Sens* 53(8):4457–4471
34. Wang J, Jiao L (2017) Application of a homogenous patch mean kernel with within-class collaborative representation for hyperspectral imagery classification. *Remote Sens Lett* 8(1):11–20

35. Wang Z, Nasrabadi NM, Huang TS (2014) Spatial-spectral classification of hyperspectral images using discriminative dictionary designed by learning vector quantization. *IEEE Trans Geosci Remote Sens* 52(8):4808–4822
36. Wu S, Chen H, Bai Y, Zhu G (2016) A remote sensing image classification method based on sparse representation. *Multimed Tools Appl* 75(19):1–18
37. Xu Y, Han Y, Hong R, Tian Q (2018) Sequential video VLAD: training the aggregation locally and temporally. *IEEE Trans Image Process* 27(10):4933–494
38. Yuan H, Feng Z, Wang M, Zhang K (2018) Self-paced learning-based probability subspace projection for hyperspectral image classification. *IEEE Trans Neural Netw Learn Syst* 99:1–6
39. Yuan H, Tang YY (2015) Sparse representation based on set-to-set distance for hyperspectral image classification. *IEEE J Selected Topics Appl Earth Observ Remote Sens* 8(6):2464–2472
40. Zhang B, Wu D, Zhang L, Jiao Q, Li Q (2012) Application of hyperspectral remote sensing for environment monitoring in mining areas. *Environ Earth Sci* 65(3):649–658
41. Zhang H, Li J, Huang Y, Zhang L (2014) A nonlocal weighted joint sparse representation classification method for hyperspectral imagery. *IEEE J Selected Topics Appl Earth Observ Remote Sens* 7(6):2056–2065
42. Zhang X, Sun Y, Jiang K, Li C, Jiao L, Zhou H (2018) Spatial sequential recurrent neural network for hyperspectral image classification. *IEEE J Selected Topics Appl Earth Observ Remote Sens* pp(99):1–15
43. Zhuang H, Deng K, Fan H, Yu M (2016) Strategies combining spectral angle mapper and change vector analysis to unsupervised change detection in multispectral images. *IEEE Geosci Remote Sens Lett* 13(5):681–685



**Chunhong Cao** received the B.S. degree in applied mathematics and software from Central South University, Changsha, China; the M.S. degree and Ph.D. degree Xiangtan University, Xiangtan, China. She is currently an Associate Professor in the College of Information Engineering, Xiangtan University. Her recent research interests include hyperspectral image processing, wavelet analysis theory, compressive sensing.



**Liu Deng** received the B.E. degree from the College of Information Engineering, Nanchang Hangkong University, Nanchang, China, in 2015. He is currently working toward the M.S. degree in Intelligent Computing and Information Processing in College of Information Engineering, Xiangtan University, Xiangtan, China. His research interests include hyperspectral image processing, pattern recognition, and machine learning.



**Wei Duan** received the B.E. degree from the school of Mathematics and Statistics, Hubei Normal University, Huangshi, China, in 2015. He is currently working toward the M.S. degree in Intelligent Computing and Information Processing in College of Information Engineering, Xiangtan University, Xiangtan, China. His research interests include hyperspectral image processing, pattern recognition, and machine learning.



**Fen Xiao** received the B.S degree, M.S degree and Ph. D degree from Xiangtan University, Xiangtan, China. She is also a visiting scholar at the State Key Laboratory of Pacific Northwest Pacific. She is currently a Professor in the College of Information Engineering, Xiangtan University. Her recent research in interests include visual saliency detection, remote sensing image analysis and Image description.



**WanChun Yang** was born in 1974. He received the Ph. D degree from Beijing University of Posts and Telecommunications. He is currently an Associate Professor in the College of Information Engineering, Xiangtan University. His research interests include mobile communication electromagnetic environment, wavelet analysis and remote sensing image analysis.



**Kai Hu** received the B.S degree, M.S degree and Ph. D degree from Xiangtan University, Xiangtan, China. He is also a visiting scholar at Nanyang Technological University. He is a member of the IEEE. His current research interests include machine learning, pattern recognition and image processing. He is a reviewer for Expert Systems with Applications, IEEE Transactions on Image Processing.