# Class-wise dictionary learning for hyperspectral image classification

**Article** *in* Neurocomputing · August 2016

DOI: 10.1016/j.neucom.2016.05.101



**4 authors**, including:

Siyuan Hao

**7** PUBLICATIONS **18** CITATIONS

SEE PROFILE

Wei Wang

Università degli Studi di Trento

**9** PUBLICATIONS **44** CITATIONS

SEE PROFILE

Lorenzo Bruzzone

Università degli Studi di Trento

**478** PUBLICATIONS **15,419** CITATIONS

SEE PROFILE

Neurocomputing 220 (2017) 121–129

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

# Class-wise dictionary learning for hyperspectral image classification

Siyuan Hao [a,*], Wei Wang [b], Yan Yan [b], Lorenzo Bruzzone [b]

[a] College of Communication and Electronic Engineering, Qingdao University of Technology, Qingdao 266520, China
[b] Department of Information Engineering and Computer Science, University of Trento, Italy

ARTICLE INFO

*Article history:*
Received 7 January 2016
Received in revised form
21 March 2016
Accepted 4 May 2016
Available online 22 August 2016

*Keywords:*
Hyperspectral image classification
Class-wise dictionary learning
Regularizer
Collaborative representation
Remote sensing

ABSTRACT

In order to effectively exploit the intra-class and inter-class structure information, we propose a new class-wise dictionary learning method for hyperspectral image classification. First, we construct two special manifold regularizers to encourage intra-class basis sharing and inter-class basis competition, and the regularizers are incorporated into the objective function to learn a discriminative class-wise dictionary. Then the sparse representations can be obtained via the learned class-wise dictionary under the collaborative representation framework. Finally, we put the sparse representations of the data into the support vector machine (SVM) for training and then apply the SVM classifiers to predict labels for the test set. The experimental results obtained on two hyperspectral datasets demonstrate that the proposed method can obtain higher classification accuracy with much lower computational cost compared with other traditional classifiers.



## 1. Introduction

Sparse representation (SR) has shown its effectiveness in many different fields of data analysis, such as computer vision [42–44] and remote sensing [1–3]. Many versions of the SR have been presented recently. For example, the orthogonal matching pursuit (OMP) [4] with smoothing and simultaneous OMP (SOMP) were proposed to incorporate the spatial information into the SR recovery problem [5]. Liu et al. integrated a spatial-spectral kernel into the SR framework [6]. One of the important properties that must be highlighted is that the SR encodes the data into discriminative sparse vectors even though the original data are quite similar to each other [7]. Because of this property, we can use the sparse representation of the data to characterize land-covers, and then employ statistical models (e.g., SVM and Deep CNN [45]) to perform the classification. In addition, if the dictionary size is small, we can use the SR of data as features for classification can effectively mitigate the effects of the Hughes phenomenon [8].

The SR with $l_0$-regularizer or $l_1$-regularizer always has a small amount of nonzero entries in the sparse vector and has very limited discriminative information. However, the sparse model with the weaker regularizer usually has more nonzero entries and it can also increase the discriminative power of the sparse vector, which promotes the introduction of the collaborative representation (CR). Different from the SR framework, the CR uses a much weaker $l_2$-regularizer, but also reduces the computation costs dramatically. Therefore, Zhang et al. [9] argued that the CR framework played a more important role than the SR framework in classification problems. In recent years, CR has been used in hyperspectral image classification tasks [10–12]. For example, Li et al. [13] proposed a nonlocal joint CR with a locally adaptive dictionary. Waqas et al. [14] developed a collaborative neighbor representation method, in which the dictionary was automatically selected. Li et al. [15] designed a kernel collaborative representation model with a Tikhonov regularizer which considered the local spatial information. Li et al. [16] presented a joint CR classification with multitask learning framework to acquire the sparse vectors and the adaptive weight for each pixel.

The SR and CR frameworks are both based on dictionary learning. Thus, their performances highly depend on the learned dictionary [17]. The characteristics of a well learnt dictionary are summarized as follows: (1) it captures the subspace structure of each class, (2) it is compact, and (3) it has a reliable generative and discriminative capability. The conventional scheme involved with hyperspectral image processing uses all the training samples as the dictionary [5]. The bases in the dictionary can be randomly selected from the training set (Random) [18], or can be constructed by K-mean clustering (K-mean) [19]. However, these schemes might be impractical for a small-sized training set and also might fail in reflecting the inter-class and intra-class distributions. Other dictionary construction schemes [20,21] are on the assumption that the samples from the same class share similar characteristics to iteratively optimize the dictionary. This is the base of learning

* Corresponding author.
*E-mail address:* lemonbananan@163.com (S. Hao).

http://dx.doi.org/10.1016/j.neucom.2016.05.101
0925-2312/

vector quantization (LVQ) and K-means-based singular value decomposition (K-SVD). In [22], a variant of the K-SVD was presented that exploited the linear extension of graph embedding to optimize the K-SVD for a small yet over complete dictionary. Some methods learn multiple dictionaries to promote discrimination between classes. Jenatton et al. [23] proposed an approach that learns structured dictionaries embedded in a hierarchy based on a tree-structured sparse regularizer to effectively exploit the semantic relations between the dictionary bases. Mairal et al. used multiple category-specific dictionaries to increase the discriminative power of different classes [24].

The additional information, such as spatial information and shared information, can be used for a better performance [35,36]. A suite of dictionary learning methods has been proposed recently, which exploits the additional information to learn a compact and yet discriminative dictionary [25–28,34,41]. Based on the hypothesis that pixels in a local region tend to have homogenous spectral behavior, several attempts have been made to incorporate the spatial information during the dictionary learning. For example, Soltani-Farani et al. presented a spatial-aware dictionary model, which is implemented by a joint sparse regularizer to induce a common sparse vector in the sparse coefficients of a contextual group [7]. Castrodad et al. designed the block-structured discriminative dictionaries, in which a spatial-spectral coherence regularizer allows pixel classification to be guided by similar neighbors [29]. Aimed at exploiting the distribution of classes, Jiang et al. added a label consistency regularizer term and a joint term (i.e., combined a label consistency term with the classification error) into the dictionary optimization objective, which are referred to as LC-KSVD1 and LC-KSVD2, respectively [30]. Winn et al. automatically constructed an appearance-based object class dictionary which considered the tradeoff between intra-class compactness and inter-class discrimination [31]. Moreover, a Fisher discriminative constraint containing the class labels was incorporated into the objective function in [32]. Guo et al. introduced the pairwise constraints and combined them with the classification error term to formulate the objective function of dictionary learning for greater discriminative power [26].

In order to learn a discriminative dictionary via the intra-class and inter-class information, we propose a novel class-wise dictionary method for hyperspectral image classification. Both the intra-class basis sharing and the inter-class basis competition regularizers are incorporated into the dictionary learning objective function, and the sparse features can then be obtained via the learned dictionary under the CR framework. Finally, we use the sparse features to train the SVM classifiers and exploit the test set for evaluation. The proposed method differs from the traditional dictionary learning it not only constructs two class-wise regularizers during the dictionary learning, but also employs the CR framework to calculate the sparse features of each pixel. In greater detail, the contributions of the proposed method are as follows:

1) Instead of retaining all the training samples in the dictionary directly, the proposed method constructs two regularizer terms (i.e., the intra-class basis sharing term and the inter-class basis competition term), which encourage the intra-class basis sharing and the inter-class basis competition;
2) Due to the limited discriminative information in the SR framework, the CR framework is employed to obtain more discriminative sparse features with much lower computational cost.

The rest of the paper is organized as follows. Section 2 briefly reviews the related work. Section 3 introduces the proposed method. Section 4 evaluates the effectiveness of the proposed method using two hyperspectral datasets. Section 5 analyzes the experimental results and concludes the paper.

## 2. Related work

### 2.1. Collaborative representation

Let $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_K]$ be a dictionary with $K$ bases. A test sample $\mathbf{z}$ can be represented as a linear combination of dictionary bases (i.e., $\mathbf{z} = \mathbf{Ds} + \varepsilon$, where $\mathbf{s} \in \mathbf{R}^K$ is a sparse vector composed of the corresponding coefficients for the bases and $\varepsilon$ is a threshold). In order to compute the sparse vector for any new input sample, $l_m$-regularizer is used in the following optimization problem:

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \|\mathbf{s}\|_m \quad s.t. \quad \|\mathbf{z} - \mathbf{Ds}\|_2^2 < \varepsilon, \tag{1}$$

when $m=1$, the obtained sparse vector is defined as the spare representation of the input sample. Similarly, it is defined as the collaborative representation when $m=2$. The biggest difference between the collaborative representation and the sparse representation is that the former operates on the concept that a test sample can be represented by a linear combination of dictionary bases of all classes, that is, samples in the $i$th class can be helpful to represent the samples in the $j$th class. On the contrary, the later only uses a few bases to represent a test sample. The non-zeros entries of sparse vector in the SR framework are much less than those in the CR framework. Therefore, the sparse representation is sparser than the collaborative representation. See the example in Fig. 1.

The sparse vector has a certain discriminative capability. Thus, the sparse vector can be treated as a feature vector for the subsequent classification. However, because there are less non-zeros entries in the sparse vector obtained by the SR framework, the sparse representation contains less discriminative information than the collaborative representation. For this reason, in this paper we use the collaborative representation. The Eq. (1) is re-expressed over $\mathbf{D}$ with $l_2$-regularizer ($m=2$):

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \|\mathbf{s}\|_2^2 \quad s.t. \quad \|\mathbf{z} - \mathbf{Ds}\|_2^2 < \varepsilon, \tag{2}$$

$\hat{\mathbf{s}}$ can be derived by solving the Lagrange function of (2):

$$\hat{\mathbf{s}} = (\mathbf{D}^T\mathbf{D} + \lambda \cdot \mathbf{I})^{-1}\mathbf{D}^T z, \tag{3}$$

where $\lambda$ is a scalar regularization parameter, which balances sparsity against classification loss.

### 2.2. Dictionary learning

The performance of the sparse representation mainly depends on the efficiency of the dictionary learning scheme. However, for the dictionary learning problem of the hyperspectral image, the conventional schemes include all the training samples in the dictionary. This might weaken the scalability of the dictionary in practical cases when the size of available training samples is small, and fail to match the structure of classes. Considering the reduced computation cost and the increased discriminative information contained in the sparse vector, we apply the CR framework rather than the traditional SR framework to learn a discriminative dictionary. The learning procedure can be formulated as the following problem:

$$<\mathbf{D}^*, \mathbf{S}^*> = \arg\min_{\mathbf{D},\mathbf{S}} \sum_{i=1}^{N'} \|\mathbf{x}_i - \mathbf{Ds}_i\|_2^2 + \lambda \|\mathbf{s}_i\|_2^2, \tag{4}$$

where $\mathbf{x}_i$ is a sample in the training set, and the number of training samples is $N'$.

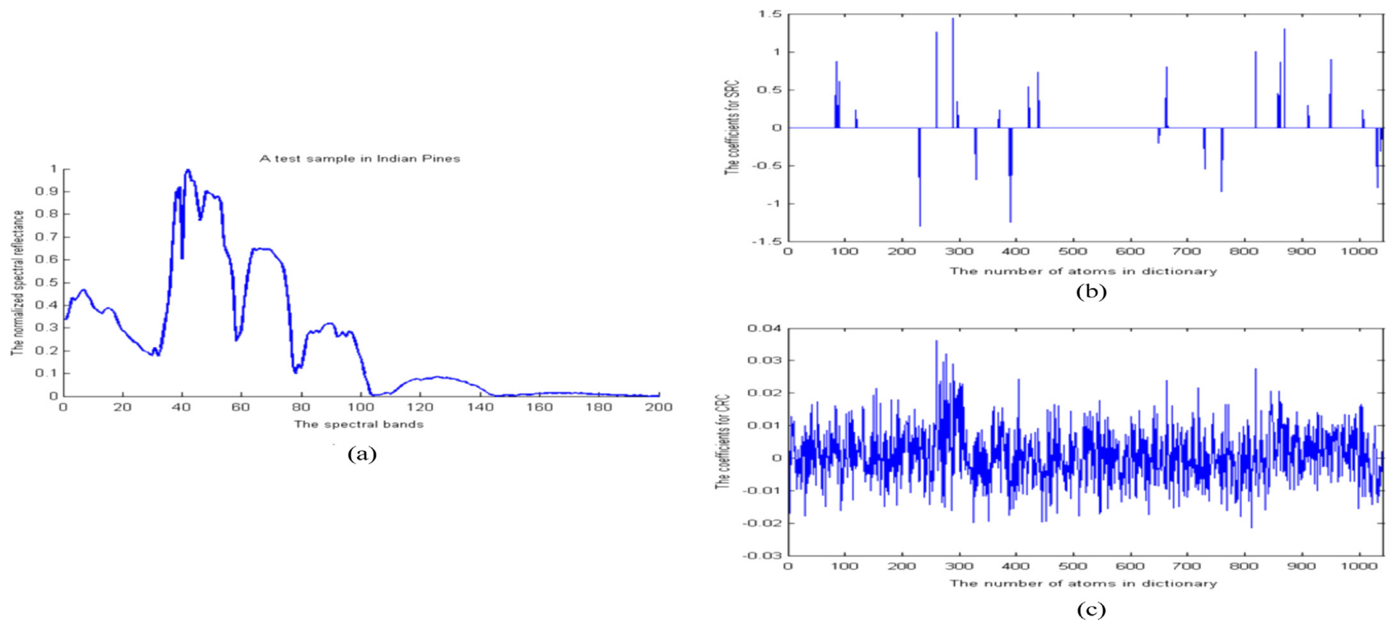



**Fig. 1.** Sparsity comparison between SR and CR: (a) A test sample in Indian Pines dataset (b) sparse vector for SR (c) sparse vector for CR. (Note: the dictionary is composed of 10% samples, and the sparse level for SR is 30).

### 2.3. Support vector machine

After obtaining the sparse feature vectors based on the process described above, the selection of the basic classifiers should be taken into consideration. Due to the good performance of dealing with non-linear problem and high dimensional data, SVM has been widely applied in hyperspectral image classification [37–39].

Given the sparse representation of training set $\mathbf{S}_{TR} = \{(\mathbf{s}_1, \mathbf{y}_1), \ldots, (\mathbf{s}_{N'}, \mathbf{y}_{N'})\}$, labeled by $\mathbf{y}_i \in \{+1, -1\}$, $i = 1, .., N'$, the optimization problem of SVM is characterized as follows:

$$\begin{aligned} &\min\ \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N'}\xi_i \\ &s.\ t.\ \ \mathbf{y}_i(\boldsymbol{w}^T\cdot\varphi(\mathbf{s}_i) + b) \geq 1 - \xi_i \\ &\xi_i \geq 0,\ \ \text{i}= 1, 2, \cdots N', \end{aligned} \tag{5}$$

where $\boldsymbol{w}$ and $b$ are the weight vector and threshold of decision function, $\xi_i$ is the slack variable measuring the degree of misclassification of $\mathbf{s}_i$, $C$ is the penalty factor, and $\varphi(\cdot)$ denotes the non-linear mapping function. The prediction function of the test sample $\mathbf{s}$ can be calculated as follows:

$$f(\mathbf{s}) = \text{sgn}\left(\sum_{i=1}^{N'}\alpha_i^{*} y_i K(\mathbf{s}, \mathbf{s}_i) + b^{*}\right), \tag{6}$$

where $\boldsymbol{\alpha}^*$ and $b^*$ are the optimal solutions of the Lagrange function of Eq. (5), and $K(\mathbf{s}, \mathbf{s}_i) = \varphi(\mathbf{s})^T\varphi(\mathbf{s}_i)$ is a kernel function.

## 3. The proposed framework

### 3.1. Intra-class basis sharing and inter-class basis competition

Conventional dictionary schemes retain all training samples in the dictionary for hyperspectral image classification. However, they might fail to reflect the structure of each class and ignore the relationship between the classes. In order to fully take advantage of the class-wise information during the dictionary learning, we propose the intra-class basis sharing and the inter-class basis competition regularizers. The theoretical foundation of the two class-wise regularizers is based on the assumptions: (1) the sparse vectors from the same class should always share similar dictionary bases, and based on this assumption, we propose a basis sharing reguarizer (2) the sparse vectors from the different classes should compete for the bases, which leads to the inter-class basis competition regularizer. The visual interpretation is shown in Fig. 2. In Fig. 2, we can clearly observe that the sparse matrix of the same class has non-zero entries in the same row, while the sparse matrixes from different classes have non-zero entries in different rows. In other words, we construct an intra-class basis sharing regularizer to compact the sparse vectors within the same class, conversely, we construct the inter-class basis competition regularizer to promote the inter-class competition, which can be achieved by making the inner product between classes smaller.

### 3.2. Formulation and optimization

Based on the assumptions above, we propose a novel class-wise dictionary method to learn an informative and discriminative dictionary, in which the two class-wise regularizers (i.e., the intra-class basis sharing regularizer and the inter-class basis competition regularizer) are integrated into the dictionary learning process. The proposed class-wise dictionary can be learnt by minimizing the following optimization problems:

$$\begin{aligned} &<\mathbf{D}^*, \mathbf{S}^*> \\ &= \arg\min_{\mathbf{D},\mathbf{S}} \sum_{i=1}^{N'}\left(\|\mathbf{x}_i - \mathbf{D}\mathbf{s}_i\|_2^2 + \lambda\|\mathbf{s}_i\|_2^2\right) \\ &\quad + \alpha\sum_{p=1}^{N}\sum_{\substack{\mathbf{s}_i\in\mathbf{S}_p \\ \mathbf{s}_j\in\mathbf{S}/\mathbf{S}_p}}(\mathbf{s}_i^T\mathbf{s}_j)^2 + \beta\sum_{i,j}^{N'}\left(\|\mathbf{s}_i - \mathbf{s}_j\|_2^2\mathbf{W}_{ij}\right), \end{aligned} \tag{7}$$

where $\mathbf{x}_i \in \mathbf{X}$ is a sample in the training set with $N'$ samples from $N$ classes, and its corresponding sparse vector is $\mathbf{s}_i$, the sparse coefficient matrix for the *pth* class is $\mathbf{S}_p$, $p \in \{1, 2, \ldots, N\}$. The last two regularizer terms in Eq. (7) are the inter-class basis competition regularizer and intra-class basis sharing regularizer, which seek bases competition between classes, while bases sharing within the same class. The parameters $\alpha \leq 1$ and $\beta \leq 1$ are the weights for each regularizer. However, the class-wise regularizers

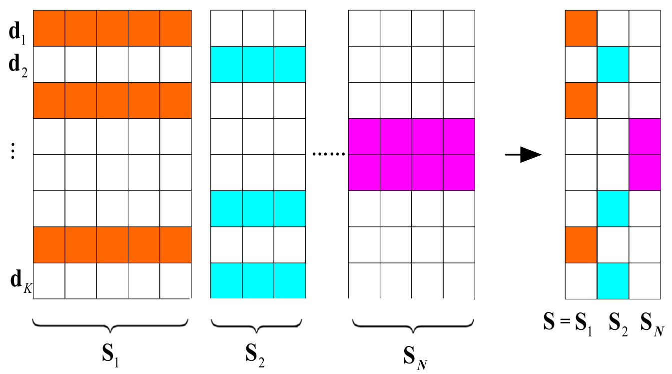



**Fig. 2.** Diagram of the intra-class basis sharing and inter-class basis competition. ($\mathbf{S}_p$, $p \in \{1, 2, \ldots, N\}$ is the sparse matrix for the *pth* class, $N$ is the number of classes, and $\mathbf{d}_i$, $i \in \{1, 2, \ldots, K\}$ represents the $i$th atom in the dictionary.).

are not the hard constraints, and the misclassification loss plays the most important role in dictionary learning. In the Eqn. above, the similarity matrix $\mathbf{W}$ can be calculated by the Eq. (8), and the similar construction can be found in the reference [40]. $\mathbf{L}$ is the Laplace graph of $\mathbf{W}$, and $\mathbf{L}_i$ is the $i$th column of $\mathbf{L}$.

$$\mathbf{W}_{ij} = \begin{cases} 1 & if\ \ \mathbf{s}_i, \mathbf{s}_j \in the\ same\ class \\ 0 & otherwise\ \ i, j \in 1, 2, \ldots, N' \end{cases}. \tag{8}$$

At the same time, it should be emphasized that the optimization problem (7) is an unconstrained non-convex with two matrix variables $\mathbf{S}$ and $\mathbf{D}$, thus, it only has local optimal solutions. However, Eq. (7) is convex with respect to any of the two variables when fixing the other one. Therefore, we update the sparse matrix $\mathbf{S}$ and dictionary $\mathbf{D}$ alternately.

When computing the sparse matrix $\mathbf{S}^*$ with the fixed D, Eq. (7) is equivalent to:

$$F(\mathbf{s}_i) = \min_{\mathbf{s}_i} \ \|\mathbf{x}_i - \mathbf{D}\mathbf{s}_i\|_2^2 + \lambda\|\mathbf{s}_i\|_2^2 + \alpha \sum_{t=1}^{N} \sum_{\substack{\mathbf{s}_i \in \mathbf{S}_t \\ \mathbf{s}_j \in \mathbf{S}/\mathbf{S}_t}} \left(\mathbf{s}_i^T \mathbf{s}_j\right)^2 + 2\beta\left(2\mathbf{s}_i^T(\mathbf{S}\mathbf{L}_i) - \mathbf{s}_i^T \mathbf{s}_i \mathbf{L}_{ii}\right). \tag{9}$$

Let the first derivative of Eq. (9) equal to zero, the updated sparse vector can be calculated:

$$\mathbf{s}_i^* = \left(\mathbf{D}^T\mathbf{D} + \lambda\mathbf{I} + \alpha \sum_{\mathbf{s}_j \in \mathbf{S}/\mathbf{S}_t} (\mathbf{s}_j\mathbf{s}_j^T) + 2\beta\mathbf{L}_{ii}\mathbf{I}\right)^{-1} \left(\mathbf{D}^T\mathbf{x}_i - 2\beta\sum_{k \neq i} \mathbf{s}_i\mathbf{L}_{ki}\right). \tag{10}$$

After computing the sparse matrix, we obtain the objective function to update the dictionary $\mathbf{D}$ with the fixed $\mathbf{S}$ as follows:

$$\min_{\mathbf{D}} \sum_{i=1}^{N'} \left(\|\mathbf{x}_i - \mathbf{D}\mathbf{s}_i\|_2^2\right., \qquad s.\,t.\ \ \|\mathbf{d}_j\|_2^2 \le \varepsilon, \quad \forall\, j = 1 \ldots K. \tag{11}$$

where $\mathbf{d}_j$ represents the $j$th atom in the dictionary, and $K$ is the number of the atoms. The optimal solution $\mathbf{D}^*$ can be computed as:

$$\mathbf{D}^* = \mathbf{X}\mathbf{S}^T\left(\mathbf{S}\mathbf{S}^T + \mathbf{I}\right)^{-1}. \tag{12}$$

The complexity of the proposed classification method can be divided into three parts: dictionary learning, sparse features computation and classification. We mainly focus on the most time-consuming part, namely, the dictionary learning part, and compare it with other dictionary learning methods from the theoretical viewpoint. LC-KSVD is an efficient dictionary method using the label information and bounded by the complexity of K-SVD, whose complexity is similar to that of the OMP [33]. OMP is a common approximation technique to solve the SR problem with $l_0$ or $l_1$-regularizer, and its implementation is quite computationally expensive in comparison with the CR problem with $l_2$-regularizer who has a closed form solution. LVQ is also built on the SR model and solved by OMP. Because the proposed method is based on the CR framework, it is an efficient approach.

Finally, the detailed steps of the proposed class-wise dictionary learning for hyperspectral image classification can be summarized as follows:

**Algorithm 1.** Proposed class-wise dictionary classification method.

Step 1) Select the training and testing sets, then initialize the original dictionary $\mathbf{D}_0$ which is composed of $t$ samples of each class from the training set, and set the involved parameters (e.g., $\lambda$, $\alpha$, $\beta$);
Step 2) Update the sparse matrix and dictionary alternately according to (10) and (12), until the change of residual in (7) is smaller than the pre-defined stopping criterion $T$;
Step 3) Calculate the sparse feature sets of the training and test sets using the CR framework over the learned dictionary;
Step 4) Train the SVM classifier, and classify the image.

## 4. Experiments

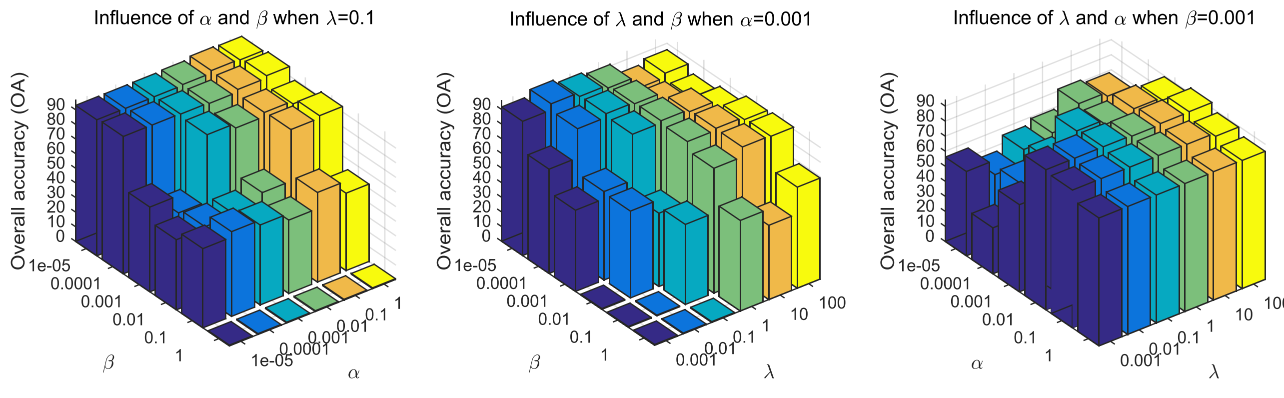
### 4.1. Datasets description and parameter setting

The experiments have been conducted on two data sets. The first hyperspectral image dataset is the Indian Pines. It contains 220 spectral bands and has a spatial size of 145 × 145 pixels, with each pixel measuring approximately 20 m by 20 m on the ground. 16 mutually exclusive classes containing from 20 to 2468 samples are considered. We randomly chose 10% of the samples as the training set and the remaining 90% as the test set.

The second hyperspectral image dataset is the University of Pavia. It mainly comprises 9 classes (i.e., asphalt, meadows, trees, metal sheets, bare soil, bitumen, bricks, shadows and gravel) with a spatial size of 610 × 340 pixels. We randomly selected 50 samples per class as training samples and the remaining ones as test samples.

In the experiments, we used the overall accuracy (OA, in percent), the average accuracy (AA, in percent) and the kappa statistic (Kappa) to quantitatively compare the performances of the tested methods. All tests were carried out with 10 trials and the averaged results were reported. The tests were implemented with MATLAB 8.2 version on a computer equipped with an Intel Core i7 Processor at 2.93-GHz and 8 GB of RAM.

In order to analyze the discriminative capability of the sparse feature, original spectral features (denoted by Original in the following description) were considered for comparison. In addition, some classical dictionary schemes were selected as the benchmarks, namely, Random, K-mean, LVQ, KSVD and LC-KSVD2. When only the intra-class basis sharing or inter-class basis competition was used, the proposed class-wise dictionary was simplified to sharing dictionary (Sharing) and competition dictionary (Competition) to illustrate the influence of the two regularizers. The parameter $\lambda$ of the CR framework varied within the range of $[10^{-3}, 10^{-2}, \ldots, 10^{2}]$. For the SVM classifiers, we used the one-against-rest multiple classification strategy with an RBF kernel, and a 10-fold cross validation was used to select the optimal penalty factor and the width of Gaussian kernel. The parameters for KSVD were set to




(a) (b) (c)

**Fig. 3.** Influence of the parameters for the Indian Pine dataset.

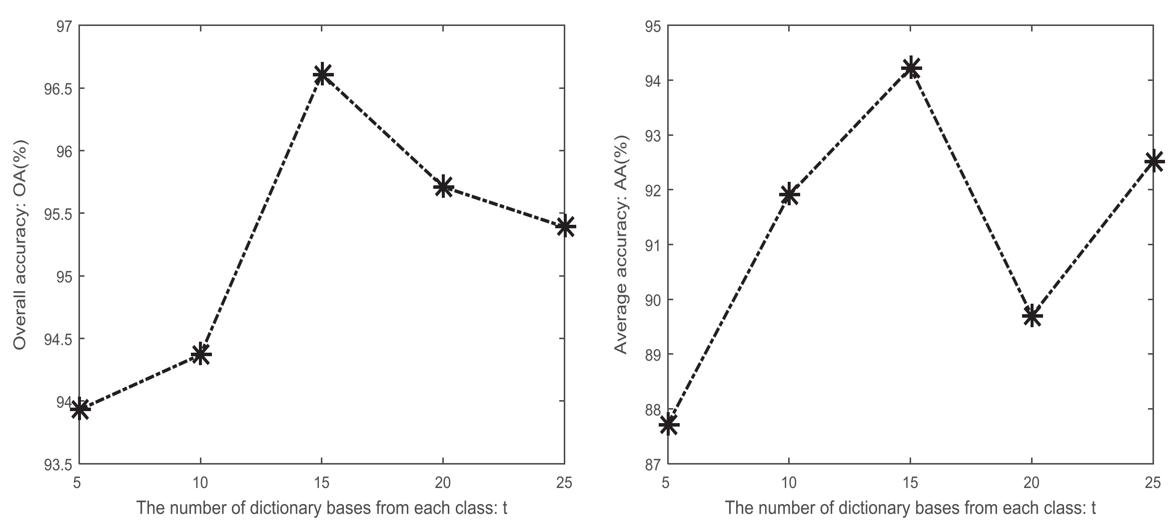
the default values reported in [20]. With respect to LC-KSVD2, the sparsity level, the weights for the label constraint term and the classification error term were set to 30, 0.01 and 0.1, respectively. For the proposed method, the weights for the inter-class basis competition regularizer and the intra-class basis sharing regularizer were both varied within the range of $[10^{-5}, 10^{-4}, \ldots, 10^{\circ}]$.

### 4.2. Analysis of the sensitivity to the parameters

In this experiment, we evaluated the sensitivity of three parameters (i.e., $\lambda$, $\alpha$, $\beta$) to find how these parameters affect the accuracy. They balanced the contributions of the reconstruction error, the inter-class basis competition and intra-class basis sharing regularizer, and varied in the ranges of $[10^{-3}, 10^{-2}, \ldots, 10^{2}]$, $[10^{-5}, 10^{-4}, \ldots, 1]$ and $[10^{-5}, 10^{-4}, \ldots, 1]$ , respectively. First, we randomly selected 15 samples per class from the training set to build the initial dictionary, and set the stopping criterion $T$=0.01. In order to better analyze the sensitivity of these parameters, we sequentially fixed one parameter each time in the following validation. A further insight into their influences for the Indian Pines dataset is provided by Fig. 3. When $\lambda$ was fixed to 0.1, we can observe in Fig. 3(a) that the proposed method is more sensitive to $\beta$ than $\alpha$. Similarly, the case with $\alpha = 0.001$ is shown in the Fig. 3(b), the sensitivity of the proposed method to $\beta$ is higher than that to $\lambda$. Moreover, Fig. 3(c) reveal that the accuracy is more sensitive to $\alpha$ than $\lambda$. In summary, the sensitivities of the proposed method to three parameters are $\beta$, $\alpha$, $\lambda$ in descending order. By comparing the different combinations of three parameters, we can see that when the parameter setting is $\lambda = 0.1$, $\alpha = 0.001$, $\beta = 0.001$, the highest accuracy can be achieved. Therefore, we fixed this parameter setting in the following tests.

### 4.3. Analysis of the influence of the dictionary size

The influence of the dictionary size on the classification performance of the proposed method was verified in this experiment. We randomly chose 10% of the samples as the training set from the Indian Pines dataset, and other parameters setting was referred to in the experiment above. $t$ denoted the number of the training samples for each class to build the dictionary, and was tuned within a set of $\{5, 10, 15, 20, 25\}$. (Note: When $t$ exceeded half of the total number of the training samples for each class, $t$ was set to half of the total number for this class). The OA vs. $t$ and AA vs. $t$ curves are plotted in Fig. 4. For Fig. 4(a), the curve increases monotonically as $t$ in the set of {5, 10, 15}, reaching the peak at $t$=15. As $t$ further increases, the curve gradually decreases. Simultaneously, when $t$ belongs to the set of {5, 10, 15}, the curve in Fig. 4(b) has the similar observation with that in Fig. 4(a). However, the following trend of AA-$t$ curve is different from that of OA-$t$ curve as $t \in \{20, 25\}$. The AA at $t$=25 has a gain of 2.82% compared to the case of $t$=20, however the OA at $t$=25 is lower than that at $t$=20, this is because the increased bases in the dictionary




(a) OA (b) AA

**Fig. 4.** Influence of the dictionary size for the Indian Pine dataset.

**Table 1**
Accuracy of the different dictionary methods for the Indian Pines dataset.

| No. | Class name | Original | Random | K-mean | LVQ | KSVD | LC-KSVD2 | sharing | competition | proposed |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Alfalfa | 77.08 | 47.92 | 14.58 | 70.83 | 62.50 | 85.42 | 87.50 | **97.92** | 95.83 |
| 2 | Corn-notill | 74.65 | 79.69 | 80.78 | 81.86 | 79.30 | 86.74 | 90.54 | 93.10 | **95.12** |
| 3 | Corn-mintill | 63.20 | 70.67 | 70.53 | 71.73 | 69.20 | 89.60 | 90.80 | **93.47** | 91.87 |
| 4 | Corn | 79.05 | 50.95 | 29.52 | 47.62 | 43.33 | 81.90 | 93.33 | **96.19** | 88.57 |
| 5 | Grass-pasture | 92.62 | 93.51 | 93.06 | 92.17 | 89.49 | 91.95 | 89.26 | 86.35 | **94.18** |
| 6 | Grass-trees | 98.21 | 93.45 | 98.36 | 94.49 | 95.54 | 99.55 | 99.40 | 99.26 | **99.85** |
| 7 | Grass-pasture-mowed | 82.61 | 30.43 | 21.74 | 21.74 | 26.09 | 91.30 | 69.57 | 86.96 | **95.65** |
| 8 | Hay-windrowed | 98.18 | 99.32 | 99.32 | 99.55 | 98.18 | 99.09 | 95.45 | 99.55 | **99.77** |
| 9 | Oats | **83.33** | 27.78 | 38.89 | 50.00 | 66.67 | 61.11 | 77.78 | 66.67 | 77.78 |
| 10 | Soybean-notill | 74.74 | 75.78 | 76.58 | 75.89 | 71.18 | 89.90 | 93.80 | **95.64** | 91.50 |
| 11 | Soybean-mintill | 85.95 | 81.77 | 78.84 | 82.76 | 83.66 | 95.32 | 96.85 | **96.94** | 95.63 |
| 12 | Soybean-clean | 85.51 | 74.09 | 86.41 | 84.60 | 79.17 | 91.12 | 92.03 | **92.75** | 91.49 |
| 13 | Wheat | 99.47 | 98.95 | 99.47 | 97.89 | 98.95 | **100** | 98.95 | 98.95 | **100** |
| 14 | Woods | 97.42 | 96.48 | 98.97 | 97.16 | 97.77 | **99.91** | **99.91** | 99.48 | 99.66 |
| 15 | Bldgs-grass-tree-drives | 59.36 | 51.46 | 50.58 | 66.96 | 53.22 | 89.18 | 88.89 | **98.54** | 95.32 |
| 16 | Stone-steel-towers | 75.29 | 81.18 | 89.41 | 88.24 | 84.71 | 97.65 | 83.53 | **98.82** | 95.29 |
| | Overall accuracy (%) | 83.69 | 81.78 | 82.07 | 83.81 | 82.06 | 93.27 | 94.45 | 95.88 | **96.61** |
| | Average accuracy (%) | 82.92 | 72.09 | 70.44 | 76.47 | 74.93 | 90.61 | 90.47 | 93.79 | **94.22** |
| | Kappa statistic ($\kappa$) | 0.8136 | 0.7918 | 0.7954 | 0.8151 | 0.7946 | 0.9233 | 0.9367 | 0.9531 | **0.9614** |

might cause the loss of class structure information. This experiment illustrates that the proposed class-wise dictionary does not require a large-sized (or over-complete) dictionary.

### *4.4. Comparisons with other classification methods*

In order to comprehensively verify the effectiveness of the proposed class-wise dictionary, we selected Random, K-mean, LVQ, KSVD, LC-KSVD2, Sharing and Competition as the reference dictionary schemes. Without loss of generality, we used SVM as the basic classifier to deal with the achieved sparse features. And the parameter setting can refer to the above experiment. Tables 1 and 2 summarized the classification accuracy of the different dictionary schemes for the Indian Pines and University of Pavia datasets, respectively, with the best accuracy in each row in bold. For the Indian Pines dataset, we have the following observations: (1) The accuracy produced by the sparse features is highly competitive with that of the original spectral features. Therefore, the sparse features have greater discriminative capability and can replace the spectral features in the classification; (2) The number of training samples for Alfalfa, Grass-pasture-mowed and Oats classes is too small to reflect the structure information of each class, thereby resulting in low accuracy for Random, K-mean, and LVQ dictionary methods. However, the accuracy of LC-KSVD2 and the proposed method are more accurate, suggesting they benefit from the inclusion of the class information and obtain a higher accuracy; (3) Using the proposed class-wise dictionary, the AA, OA, and Kappa reach the maximum values (i.e., 94.22%, 96.61%, and 0.9614, with gains of 3.61%, 3.34%, and 0.0381, respectively, over the LC-KSVD2 dictionary method); (4) When we only added the intra-class basis sharing or inter-class basis competition regularizer into the objective function, the accuracy can also achieve a certain degree of improvement. However, the Sharing has an inferior performance than Competition. Moreover, the proposed method by the use of two regularizers leads to the optimal result. The visual comparison shown in Fig. 5 evidently validates the advantage of the proposed method, for example Bldgs-grass-tree-drives region, some pixels are seriously misclassified by Original, Random, K-mean, LVQ and KSVD, however, this region classified by LC-KSVD2 and the proposed method is relatively clear. Furthermore, the proposed method produces more accurate maps than LC-KSVD2, especially for the regions of Corn, Corn-notill and Alfalfa.

As listed in Table 2, it can be observed that the proposed method has a significant improvement over other classifiers for the University of Pavia dataset, in term of the accuracy for each class and the overall accuracy. For the dictionary schemes based on the class information, the proposed method has the improvements of 9.9% in OA, 9.49% in AA, and 0.1292 in Kappa compared to LC-KSVD2. Moreover, when only the intra-class basis sharing is used, the higher accuracy can be obtained. By analyzing the classification results in Fig. 6, the proposed method produces more accurate maps than the other methods, especially for the regions of Meadows and Bare soil. To summarize, the proposed class-wise dictionary classification method can fully use the class information to aid the classification task and has excellent performance.

**Table 2**
Accuracy of the different dictionary methods for the University of Pavia dataset.

| No. | Class name | Original | Random | K-mean | LVD | KSVD | LC-KSVD2 | Sharing | Competition | proposed |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Asphalt | 64.34 | 70.66 | 74.88 | 76.55 | 75.43 | 82.28 | **92.63** | 87.94 | 88.74 |
| 2 | Meadows | 75.08 | 76.32 | 84.09 | 82.12 | 77.45 | 87.46 | 92.24 | 91.17 | **95.24** |
| 3 | Gravel | 73.40 | 69.89 | 80.23 | 71.55 | 71.94 | 80.82 | 89.17 | 92.44 | **96.39** |
| 4 | Trees | 96.02 | 95.39 | 87.92 | 93.60 | 91.97 | 89.78 | 96.45 | 96.58 | **97.81** |
| 5 | Metal sheet | 99.61 | 99.85 | 99.61 | 99.00 | **99.92** | **99.92** | 99.61 | 99.38 | **99.92** |
| 6 | Bare soil | 71.18 | 80.40 | 82.91 | 84.29 | 82.37 | 80.96 | 99.02 | 98.71 | **99.46** |
| 7 | Bitumen | 86.41 | 87.97 | 91.80 | 90.00 | 81.64 | 90.08 | 98.98 | 97.81 | **99.30** |
| 8 | Brick | 68.23 | 83.67 | 69.88 | 76.60 | 74.20 | 75.83 | 91.00 | 92.10 | **95.62** |
| 9 | Shadows | 97.21 | 99.22 | **100** | 99.33 | 99.89 | 99.89 | 99.89 | **100** | **100** |
| | Overall accuracy (%) | 75.34 | 79.15 | 82.43 | 82.46 | 79.49 | 85.46 | 93.73 | 92.72 | **95.36** |
| | Average accuracy (%) | 81.28 | 84.82 | 85.70 | 85.89 | 83.87 | 87.45 | 95.44 | 95.13 | **96.94** |
| | Kappa statistic ($\kappa$) | 0.6849 | 0.7344 | 0.7719 | 0.7738 | 0.7384 | 0.8098 | 0.9180 | 0.9050 | **0.9390** |

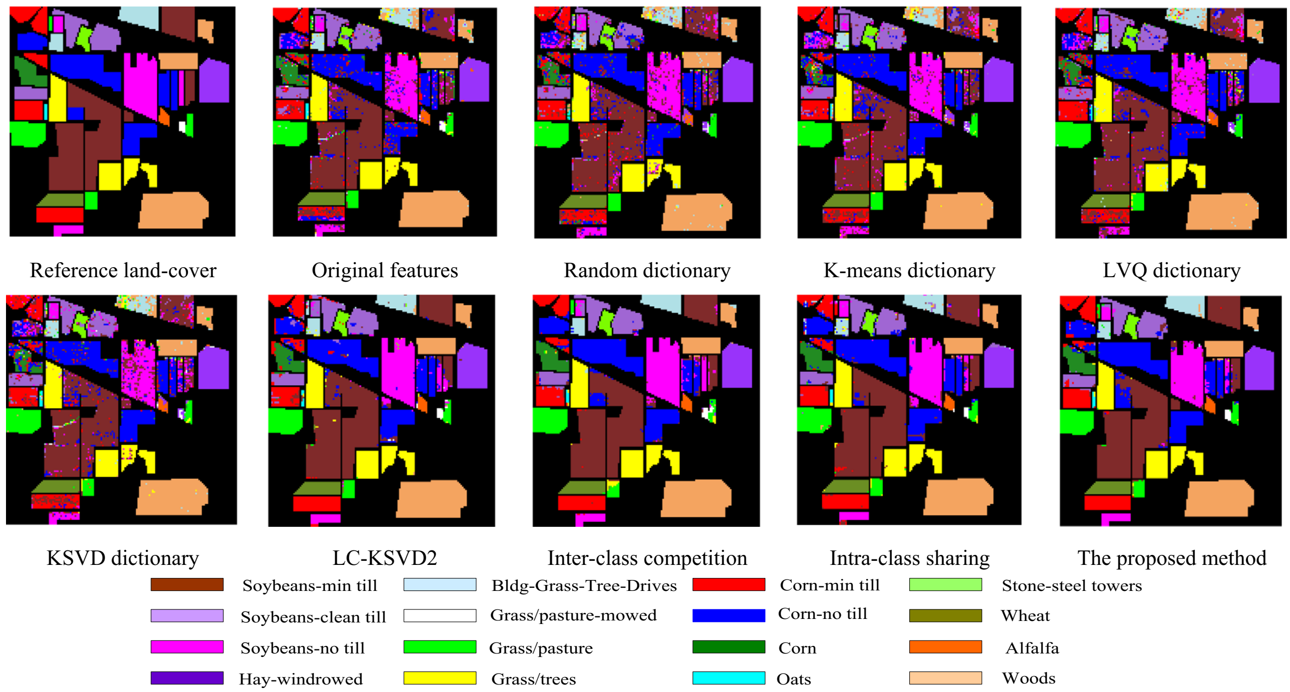



**Fig. 5.** The classification results for the Indian Pine dataset.

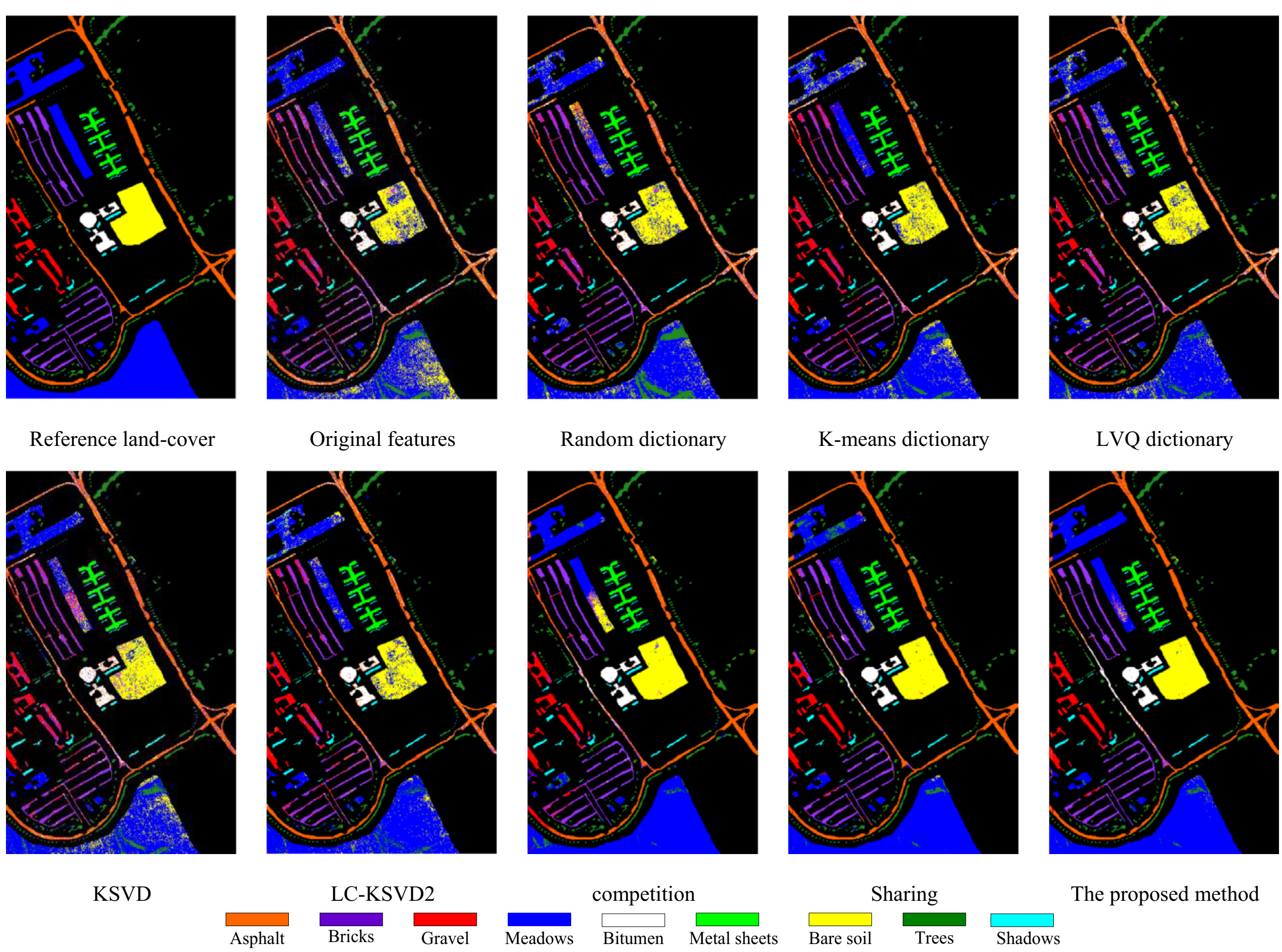



**Fig. 6.** The classification results for the University of Pavia dataset.

**Table 3**
Computational cost of different dictionary methods for the Indian Pines dataset.

| | Random | K-mean | LVQ | KSVD | LC-KSVD2 | Proposed |
|---|---|---|---|---|---|---|
| time (s) | 0.09 | 4.00 | 125.85 | 92.38 | 95.40 | 62.38 |

**Table 4**
Computational cost of different dictionary methods for the University of Pavia dataset.

| | Random | K-mean | LVQ | KSVD | LC-KSVD2 | Proposed |
|---|---|---|---|---|---|---|
| time (s) | 0.10 | 4.50 | 140.82 | 95.87 | 108.92 | 68.98 |

### 4.5. Analysis of computational cost

The running times of the different dictionary methods for the Indian Pines dataset and University of Pavia dataset are listed in Tables 3 and 4. Similar conclusions can be observed in these two tables, therefore, we only analyze Table 3 in this part. We can obtain a general rank of these dictionary methods in terms of the running time in an ascending order: Random, K-mean, the proposed class-wise dictionary, KSVD, LC-KSVD2 and LVQ. The computing time required by the Random and K-mean dictionary methods are very short, this is because their dictionary constructions do not involve any function optimization process. LVQ, KSVD and LC-KSVD2 are all based on the SR framework solved by OMP, which results in relatively long time. In addition, the proposed class-wise dictionary took about 62.38 s, which is less than those of LVQ, KSVD and LC-KSVD. The main reason is that the proposed class-wise dictionary operates on the CR framework, which uses much weaker $l_2$-regularizer and the optimized dictionary can be calculated in a closed form. Therefore, the proposed class-wise dictionary is proved to be an efficient approach to learn a discriminative dictionary.

## 5. Discussion and conclusion

In this paper, we proposed a novel class-wise dictionary method for hyperspectral image classification, which is composed of three main parts: dictionary learning, sparse features calculation and classification. The novelty is the introduction of the two regularizers (i.e., intra-class basis sharing and inter-class basis competition), which seek bases sharing within the same class, while resist bases sharing between classes. Experiments on two hyperspectral image datasets were conducted to validate the effectiveness and advantages of the proposed method: the sensitivity analysis, the influence of the parameters analysis, comparisons with others classifiers and the computational cost analysis. The experimental results reveal that the biggest advantage of the proposed dictionary method over the traditional ones is that the class-wise information which reflects class structure is taken into consideration to improve the discriminative capability of bases in the dictionary. Another advantage is the relatively lower computational cost.

Furthermore, it is worthwhile to discuss the problem found in the experiments. Comparing Table 1 with Table 2, the Competition is more suitable for the Indian Pines dataset with respect to the Sharing, with 95.88% in OA, 93.79% in AA, and 0.9531 in Kappa. However, the Sharing trends to be more advantageous for the University of Pavia observed in Table 2. Therefore, we cannot determine which regularizer is the best one, but when two regulariers are involved in the dictionary learning, it tends to produce higher classification accuracy. Thus, in future research, more cases need to be tested to explore the contribution of each regularizer. In addition, dictionary learning is a classical unsupervised learning method. It would be also worthwhile to incorporate the supervised information and the structure of each class into the dictionary optimization process.

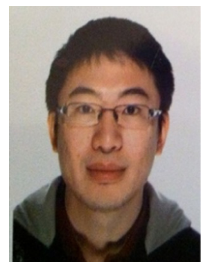
## References


[1] J. Yang, K. Yu, Y. Gong, and T. Huang, Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification, Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2009.
[2] A.S. Charles, B.A. Olshausen, C.J. Rozell, Learning Sparse Codes for hyperspectral Imagery, IEEE J. Sel. Top. Signal Process. 5 (5) (2011) 963–978.
[3] L. Fang, S. Li, X. Kang, J.A. Benediktsson, Spectral-spatial Hyperspectral Image Classification via Multiscale Adaptive Sparse Representation, IEEE Trans. Geosci. Remote Sens. 52 (12) (2014) 7738–7749.
[4] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition, Proc. of the 27$^{th}$ Annual Asilomar Conference on Signals Systems and Computers, vol. 1, pp. 40-44, 1993.
[5] Y. Chen, N.M. Nasrabadi, T.D. Tran, Hyperspectral image classification using dictionary-based sparse representation, IEEE Trans. Geosci. Remote Sens. 49 (10) (2011) 3973–3985.
[6] J. Liu, Z. Wu, Z. Wei, L. Xiao, L. Sun, Spatial-Spectral Kernel Sparse Representation for Hyperspectral Image Classification, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 6 (6) (2013) 2462–2471.
[7] A. Soltani-Farani, H.R. Rabiee, S.A. Hosseini, Spatial-aware dictionary learning for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 53 (1) (2015) 527–541.
[8] B.M. Shahshahani, D. Landgrebe, The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon, IEEE Trans. Geosci. Remote Sens 32 (5) (1994) 1087–1095.
[9] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" In Proc. Int. Conf. Comput. Vis., Barcelona, Spain, Nov. 2011, pp. 471-478.
[10] J. Li, H. Zhang, Y. Huang, L. Zhang, Hyperspectral image classification by nonlocal joint collaborative representation with a locally adaptive dictionary, IEEE Trans. Geosci. Remote Sens 52 (6) (2014) 3707–3719.
[11] S. Jia, L. Shen, Q. Li, Gabor Feature-Based Collaborative Representation for Hyperspectral Imagery Classification, IEEE Trans. Geosci. Remote Sens 53 (2) (2015) 1118–1129.
[12] Y. Chi, F. Porikli, Classification and Boosting with Multiple Collaborative Representations, IEEE Trans. Pattern Anal. Mach. Intell. 36 (8) (2014) 1519–1531.
[13] S. Jia, L. Deng, L. Shen, "An effective collaborative representation algorithm for hyperspectral image classification," Multimedia and Expo (ICME), 2014 IEEE International Conference on, Chengdu, China, pp.1-6, 14-18 Jul. 2014.
[14] Jadoon Waqas, Z. Yi, L. Zhang, Collaborative neighbor representation based classification using l2-minimization approach, Pattern Recognit. Lett. 34 (2013) 201–208.
[15] W. Li, Q. Du, M.M. Xiong, Kernel collaborative representation with Tikhonov regularization for hyperspectral image classification, IEEE Geosci. Remote Sens. Lett. 12 (1) (2015) 48–52.
[16] J. Li, H. Zhang, L. Zhang, X. Huang, Joint Collaborative Representation with multitask learning for Hyperspectral Image Classification, IEEE Trans. Geosci. Remote Sens. 52 (2014) 5923–5936.
[17] J. Mairal, F. Bach, J. Ponce, G. Sapiro, Online dictionary learning for sparse coding, ICML, Montr., Can. (2009) 689–696.
[18] J. Wright, M. Yang, A. Ganesh, S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009) 210–227.
[19] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, " Locality-constrained Linear Coding for Image Classification," Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2010.
[20] Michal Aharon, Michael Elad, Alfred Bruckstein, K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation, IEEE Trans. Signal Process. 54 (n11) (2006) 4311–4322.
[21] Y. Chen, N.M. Nasrabadi, and T.D. Tran, "Discriminative dictionary design using LVQ for hyperspectral image classification". Hyperspectral Image and Signal Processing (WHISPERS), 2012 4th Workshop on, Shanghai, China, pp.1-4, (4-7 Jun. 2012).
[22] R. Ptucha, A.E. Savakis, LGE-KSVD: robust sparse representation classification, IEEE Trans. Image Process 23 (4) (2014) 1737–1750.
[23] R. Jenatton, J. Mairal, G. Obozinski, and F. Bach, "Proximal Methods for Sparse Hierarchical Dictionary Learning," Proc. International Conference on Machine Learning, 2010.
[24] J. Mairal, M. Leordeanu, F. Bach, M. Hebert, and J. Ponce, "Discriminative Sparse Image Models for Class-Specific Edge Detection and Image Interpretation," Proc. European Conf. Computer Vision, 2008.
[25] Z. Wang, N.M. Nasrabadi, T.S. Huang, Spatial–spectral classification of hyperspectral images using discriminative dictionary designed by learning vector quantization, IEEE Trans. Geosci. Remote Sens 52 (8) (2014) 4808–4822.
[26] H. Guo, Z. Jiang, Larry S. Davis, "Discriminative dictionary learning with

pairwise constraints," Proceedings of the 11th Asian conference on Computer Vision, 2012.
[27] N. Zhou, Y. Shen, J. Peng, and J. Fan, "Learning Inter-related Visual Dictionary for Object Recognition," Pro. IEEE Conf. Computer Vision and Pattern Recognition, 2012.
[28] Z. Jiang, G. Zhang, and L. Davis, "Submodular Dictionary Learning for Sparse Coding," Pro. IEEE Conf. Computer Vision and Pattern Recognition, 2012.
[29] A. Castrodad, X. Zhengming, J. Greer, E. Bosch, L. Carin, G. Sapiro, Learning discriminative sparse reprensentations for modeling, source separation, and mapping of hyperspectral imagery, IEEE Trans. Geosci. Remote Sens. 49 (11) (2011) 4263–4281.
[30] Z. Jiang, Z. Lin, L.S. Davis, Label consistent K-SVD: Learning a discriminative dictionary for recognition, IEEE Trans. Pattern Anal. Mach. Intell. 35 (11) (2013) 2651–2664.
[31] J. Winn, A. Criminisi, and T. Minka, "Object Categorization by Learning Universal Visual Dictionary," Proc. IEEE International Conference on Computer Vision, 2005.
[32] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Fisher Discrimination Dictionary Learning for Sparse Representation," Pro. IEEE International Conference on Computer Vision, 2011.
[33] R. Rubinstein, M. Zibulevsky, and M. Elad, "Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit," Technical Report-CS Technion., 2008.
[34] W. Wang, Y. Yan and N. Sebe, "Attribute Guided Dictionary Learning," Proceedings of International Conference on Multimedia Retrieval, Shanghai, China, 2015, pp. 211-218.
[35] Zhigang Ma, Yi Yang, Nicu Sebe, Alexander G. Hauptmann, Knowledge Adaptation with Partially Shared Features for Event Detection Using Few Exemplars, IEEE Trans. Pattern Anal. Mach. Intell. 36 (9) (2014) 1789–1802.
[36] Y. Yang, Z. Ma, A.G. Hauptmann, et al., Feature Selection for Multimedia Analysis by Sharing Information Among Multiple Tasks, IEEE Trans. Multimed. 15 (3) (2013) 661–669.
[37] Biting Yu, Bo Jia, Lu Ding, Zhengxiang Cai, et al., Hybrid dual-tree complex wavelet transform and support vector machine for digital multi-focus image fusion, Neurocomputing 182 (2016) 1–9.
[38] Lingxi Xie, Jingdong Wang, Bo Zhang, Qi Tian, Incorporating visual adjectives for image classification, Neurocomputing 182 (2016) 48–55.
[39] Douglas O. Cardoso, Danilo S. Carvalho, Daniel S.F. Alves, et al., Financial credit analysis via a clustering weightless neural classifier, Neurocomputing 183 (2016) 70–78.
[40] Yi Yang, Feiping Nie, Dong Xu, Jiebo Luo, Yueting Zhuang, Yunhe Pan, A Multimedia Retrieval Framework Based on Semi-Supervised Ranking and Relevance Feedback, IEEE Trans. Pattern Anal. Mach. Intell. 34 (4) (2012) 723–742.
[41] Xiaojun Chang, Yi Yang, Alexander Hauptmann, Eric P Xing, Yao-Liang Yu, "Semantic Concept Discovery for Large-Scale Zero-Shot Event Detection," AAAI Publications, Twenty-Fourth International Joint Conference on Artificial Intelligence, 2015.
[42] Mohammad Akbari, Liqiang Nie, Tat-Seng Chua, aMM: Towards adaptive ranking of multi-modal documents, Int. J. Multimed. Inf. Retr. 4 (4) (2015) 233–245.
[43] J. Shen, R.H. Deng, Z. Cheng, et al., On Robust Image Spam Filtering via Comprehensive Visual Modeling, Pattern Recognit. 48 (10) (2015) 3227–3238.
[44] Nie L, Yan S, Wang M, et al., "Harvesting visual concepts for image search with complex queries," Proceedings of the 20th ACM international conference on Multimedia. ACM, 2012, pp. 59-68.
[45] Y Wei, W Xia, J Huang, et al. "CNN: Single-label to Multi-label," Eprint Arxiv, 2014.

**Siyuan Hao** received her pH.D. degree from the College of Information and Communications Engineering in Harbin Engineering University, Harbin, China, in 2015. She is currently a Researcher at Qingdao University of Technology, China, where she teaches remote sensing and electrical communication.Her research interests focus on hyperspectral imagery processing and machine learning.

**Wei Wang** received the master degree from the University of Southern Denmark. He is currently a PhD student in Multimedia and Human Understanding Group in the University of Trento, Italy. His research interests include machine learning and its application to computer vision and multimedia analysis.



**Yan Yan** received the pH.D. degree from the University of Trento, Trento, Italy, in 2014, where he is currently a Post-Doctoral Researcher with the Multimedia and Human Understanding Group. His research interests include machine learning and its application to computer vision and multimedia analysis.

**Lorenzo Bruzzone** (S'95–M'98-SM'03–F'10) received the "Laurea" (M.S.) degree in electronic engineering (summacum laude) and the pH.D. degree in telecommunications from the University of Genoa, Italy, in 1993 and 1998, respectively. He is currently a Full Professor of telecommunications at the University of Trento, Italy, where he teaches remote sensing, radar, pattern recognition, and electrical communications. Dr. Bruzzone is the founder and the director of the Remote Sensing Laboratory in the Department of Information Engineering and Computer Science, University of Trento. His current research interests are in the areas of remote sensing, radar and SAR, signal processing, and pattern recognition. He promotes and supervises research on these topics within the frameworks of many national and international projects. Among the others, he is the Principal Investigator of the Radar for icy Moon exploration (RIME) instrument in the framework of the JUpiter ICy moons Explorer (JUICE) mission of the European Space Agency. He is the author (or coauthor) of 161 papers in referred international journals (111 in IEEE journals), more than 220 papers in conference proceedings, and 17 book chapters. He is editor/co-editor of 16 books/conference proceedings. His papers are highly cited, as proven form the total number of citations (more than 12900) and the value of the h-index (58) (source: Google Scholar). He was invited as keynote speaker in more than 30 international conferences and workshops. Since 2009 he is a member of the Administrative Committee of the IEEE Geoscience and Remote Sensing Society (GRSS). Dr. Bruzzone ranked first place in the Student Prize Paper Competition of the 1998 IEEE International Geoscience and Remote Sensing Symposium (Seattle, July 1998). Since that time he was recipient of many international and national honors and awards, including the recently achieved IEEE GRSS 2015 Outstanding Service Award. Dr. Bruzzone was a Guest Co-Editor of different Special Issues of international journals. He is the co-founder of the IEEE International Workshop on the Analysis of Multi-Temporal Remote-Sensing Images (MultiTemp) series and is currently a member of the Permanent Steering Committee of this series of workshops. Since 2003 he has been the Chair of the SPIE Conference on Image and Signal Processing for Remote Sensing. Since 2013 he has been the founder Editor-in-Chief of the IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE. Currently he is an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and the CANADIAN JOURNAL OF REMOTE SENSING. Since 2012 he has been appointed Distinguished Speaker of the IEEE Geoscience and Remote Sensing Society.