Spectral–Spatial Classification of Hyperspectral Images With a Superpixel-Based Discriminative Sparse Model

Leyuan Fang, *Member, IEEE*, Shutao Li, *Member, IEEE*, Xudong Kang, *Student Member, IEEE*, and Jón Atli Benediktsson, *Fellow, IEEE*

***Abstract*—A novel superpixel-based discriminative sparse model (SBDSM) for spectral–spatial classification of hyperspectral images (HSIs) is proposed. Here, a superpixel in a HSI is considered as a small spatial region whose size and shape can be adaptively adjusted for different spatial structures. In the proposed approach, the SBDSM first clusters the HSI into many superpixels using an efficient oversegmentation method. Then, pixels within each superpixel are jointly represented by a set of common atoms from a dictionary via a joint sparse regularization. The recovered sparse coefficients are utilized to determine the class label of the superpixel. In addition, instead of directly using a large number of sampled pixels as dictionary atoms, the SBDSM applies a discriminative K-SVD learning algorithm to simultaneously train a compact representation dictionary, as well as a discriminative classifier. Furthermore, by utilizing the class label information of training pixels and dictionary atoms, a class-labeled orthogonal matching pursuit is proposed to accelerate the K-SVD algorithm while still enforcing high discriminability on sparse coefficients when training the classifier. Experimental results on four real HSI datasets demonstrate the superiority of the proposed SBDSM algorithm over several well-known classification approaches in terms of both classification accuracies and computational speed.**

***Index Terms*—Dictionary learning, discriminative sparse model, hyperspectral image (HSI) classification, sparse representation, superpixel.**

I. Introduction

HYPERSPECTRAL sensors can acquire digital images in more than 100 narrow contiguous bands, spanning the visible-to-infrared spectrum. Each pixel in a hyperspectral image (HSI) is a high-dimensional vector whose entries are the spectral responses of various spectral bands. The very informative spectral information of the HSI pixels can be utilized to distinguish objects in the image scene.

Manuscript received March 23, 2014; revised October 21, 2014; accepted January 11, 2015. This work was supported in part by the National Natural Science Foundation of China under Grant 61172161, by the National Natural Science Foundation for Distinguished Young Scholars of China under Grant 61325007, by the Fundamental Research Funds for the Central Universities, Hunan University, and by the Scholarship Award for Excellent Doctoral Student granted by the Chinese Ministry of Education.

L. Fang, S. Li, and X. Kang are with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China (e-mail: fangleyuan@gmail.com; shutao_li@hnu.edu.cn; xudong_kang@hnu.edu.cn).

J. A. Benediktsson is with the Faculty of Electrical and Computer Engineering, University of Iceland, 101 Reykjavík, Iceland (e-mail: benedikt@hi.is).

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

Digital Object Identifier 10.1109/TGRS.2015.2392755

In HSI classification, each pixel is labeled as belonging to one of the classes based on their spectral characteristics, given a representative training set from each class. During the last three decades, a large number of techniques have been developed for HSI classification. Among these, the support vector machine (SVM) [1] is a very powerful HSI classifier [2], [3]. Multinomial logistic regression [4]–[6], which uses the logistic function to compute the posterior probability, is another widely used classifier. The artificial immune network and biological deoxyribonucleic acid computing are two computationally intelligent models, which are also applied in HSI classification [7], [8]. Generally, although the aforementioned classifiers can effectively utilize the spectral information of the HSI, they do not consider the spatial context. Recently, to further improve classification performance, approaches based on composite kernels [9], support vector conditional random fields [10], and segmentation [11], [12], have been proposed to incorporate spatial information into the analysis of HSIs. In addition, some other recent classification approaches have focused on the design of effective feature extraction techniques (e.g., clonal selection feature-selection [13], extended morphological profiles [14], tensor discriminative locality alignment [15] and multiple features combination [16], [17]).

Sparse representation [18] has been demonstrated to be a very powerful tool for many computer vision tasks [19]–[26], often leading to state-of-the-art performance. Recently, sparse representation has been also applied in HSI classification [17], [27]–[37], using the observation that hyperspectral pixels approximately lie in a low-dimensional subspace spanned by dictionary atoms from the same class. Therefore, an unknown test pixel (i.e., a pixel that is to be classified) can be sparsely represented by a linear combination of a few atoms from the entire training dictionary. The corresponding sparse coefficients represent the positions of selected atoms and related weight values, which can be used to determine the class label of the test pixel. To further exploit the spatial contexts of the HSI, [27]–[29] define a fixed-size local region for each test pixel and then sparsely represent neighboring pixels in such a region for classification of the test pixel. This region-based sparse representation technique can provide promising classification accuracies. However, the spatial information of the HSI may not be sufficiently utilized, if the shape and the size of the region are fixed. That is, the shape of the regions should be varied according to the different spatial structures in the HSI. In addition, to classify

each pixel, many of its neighboring pixels within the local region are required to be coded, which will create a significant computational burden.

In this paper, a superpixel-based discriminative sparse model (SBDSM) is proposed to effectively exploit the spatial information of the HSI. Each superpixel in the HSI can be regarded as a small spatial region whose shape and size can be adaptively changed for different spatial structures [38]–[40]. The SBDSM first employs an efficient oversegmentation approach [40] to cluster the HSI into many superpixels. Then, pixels in each superpixel are assumed to have very similar spectral characteristics, and their correlations are exploited via a joint sparse regularization [41]. Specifically, the joint sparse regularization simultaneously represents the pixels within each superpixel using linear combinations of a few common atoms from a structured dictionary. Finally, instead of classifying each single pixel, the recovered sparse coefficients can be jointly utilized to determine the class label of the whole superpixel, and that reduces the computational cost. Note that, in some very recent works [42], [43], the HSI is also classified based on the superpixel. However, these works [42], [43] use histogram descriptors or K-means clustering for the classification of each superpixel and do not consider the correlation of pixels within each superpixel. In contrast, the proposed SBDSM can effectively capture the correlations of pixels in each superpixel by utilizing the joint sparse model.

A fundamental consideration in employing the sparse representation is the design of the structural dictionary [44]. In [27]–[29], training samples that are directly extracted from the HSI are used to form a dictionary. Usually, a dictionary constructed in the aforementioned way may not be robust enough to represent each test pixel. Furthermore, such a dictionary is of a huge size and its use in analysis creates a great computational burden. Therefore, to make the dictionary compact, representative and discriminative, the SBDSM can apply the discriminative K-SVD algorithm [45] to simultaneously learn one dictionary and one classifier. However, the K-SVD [46] is also known to be a very computationally intensive algorithm. K-SVD consists of two iterative steps, i.e., sparse coding with orthogonal matching pursuit (OMP) [47] and dictionary updating with singular value decomposition (SVD). The main computational cost is caused by the OMP algorithm [48], which requires a search of all the dictionary atoms in every iteration to pursue the sparse coefficients. In addition, the sparse coefficients obtained by the OMP algorithm lack the discriminability for appropriate training of a simple linear classifier [49], [50].

To address the aforementioned issues, motivated by works in [49] and [50], a class-labeled OMP algorithm is proposed here by incorporating the class information of the training samples and dictionary atoms into the original OMP algorithm. Since each training sample is ideally represented by atoms from the same class, the class-labeled OMP algorithm only searches atoms associated with the class of the processed training sample, greatly reducing the computational cost of sparse coding. Accordingly, the class-labeled OMP algorithm also forces training samples from a particular class to be represented by dictionary atoms from the same class. That enables training samples from the same class to have very similar sparse coefficients and, thus, enhances their discriminability for training the classifier. Note that, very recently, Wang *et al.* also introduced a dictionary learning algorithm for HSI classification [35]. In [35], Wang *et al.* learned the dictionary by utilizing a learning vector quantization technique, instead of the discriminative K-SVD framework. Thus, the learning method of Wang *et al.* is different from the proposed dictionary learning approach.

The main contributions of this paper can be summarized as follows: 1) A superpixel-based sparse model that can adaptively exploit the spatial contexts of a HSI is proposed; 2) a reduction in computational cost is obtained with a superpixel-based classification strategy; and 3) a class-labeled OMP algorithm is proposed for accelerating the dictionary learning process while enforcing high discriminability on sparse coefficients for training a classifier.

The rest of this paper is organized as follows. In Section II, the sparse representation-based HSI classification and the discriminative K-SVD dictionary learning algorithm are briefly reviewed. Section III introduces the proposed SBDSM framework for HSI classification. Experimental results on four images are presented in Section IV. Section V concludes this paper and suggests future research directions.

II. Sparse Representation-Based HSI Classification and Discriminative K-SVD Dictionary Learning Algorithm

A. Sparse Representation-Based HSI Classification

We denote one pixel of HSI as $\mathbf{y} \in \mathbb{R}^{S\times 1}$ with S indicating the number of spectral bands and a structural dictionary as $\mathbf{D} = [\mathbf{D}_1, \ldots, \mathbf{D}_r, \ldots, \mathbf{D}_R] \in \mathbb{R}^{S\times N}$, where $\mathbf{D}_r \in \mathbb{R}^{S\times N_r}, r = 1, \ldots, R$, is the rth class subdictionary whose columns (atoms) are directly drawn or trained from the training pixels, R is the number of classes, N_r is the number of atoms in subdictionary $\mathbf{D}_r$ and $N = \sum_{r=1}^{R} N_r$ is the total number of atoms in $\mathbf{D}$. The sparse representation-based classification (SRC) technique was first proposed in [19] for face recognition. In [27], Chen *et al.* extended the SRC to HSI classification based on the observation that the spectral pixels approximately lie in a low-dimensional subspace spanned by atoms from the same class. Therefore, a pixel $\mathbf{y}^{\text{test}}$, whose class identity is unknown, can be represented as a linear combination of atoms from all classes

$$\mathbf{y}^{\text{test}} = \mathbf{D}\alpha^{\text{test}} \tag{1}$$

where $\boldsymbol{\alpha}^{\text{test}} \in \mathbb{R}^N$ is the sparse coefficient for the test spectral pixel $\mathbf{y}^{\text{test}}$. Given the structural dictionary $\mathbf{D}$, the sparse coefficient $\boldsymbol{\alpha}^{\text{test}}$ can be obtained by solving the following error constrained problem:

$$\hat{\boldsymbol{\alpha}}^{\text{test}} = \arg\min \left\|\boldsymbol{\alpha}^{\text{test}}\right\|_0 \text{ subject to } \left\|\mathbf{y}^{\text{test}} - \mathbf{D}\alpha^{\text{test}}\right\|_2 \leqslant \sigma \tag{2}$$

or the sparsity constrained problem

$$\hat{\boldsymbol{\alpha}}^{\text{test}} = \arg\min \left\|\mathbf{y}^{\text{test}} - \mathbf{D}\alpha^{\text{test}}\right\|_2 \text{ subject to } \left\|\boldsymbol{\alpha}^{\text{test}}\right\|_0 \leqslant K \tag{3}$$

where $\|\cdot\|_0$ and $\|\cdot\|_2$ represent ℓ_0 and ℓ_2-norms, respectively, σ is the error tolerance and K is the sparsity level,

representing the number of selected atoms in the dictionary (also corresponding to the nonzeros coefficients in $\hat{\boldsymbol{\alpha}}^{\text{test}}$). The aforementioned problems are, in general, known to be nondeterministic polynomial-time hard (NP-hard) [51]. However, they can be approximately solved with OMP [47]. After the sparse coefficient vector $\hat{\boldsymbol{\alpha}}^{\text{test}}$ is obtained, we can determine the class label r^{test} of the test pixel $\mathbf{y}^{\text{test}}$ basing on the minimal representation error criteria

$$\begin{aligned}\hat{r}^{\text{test}} &= \arg\min_{r} \left\|\mathbf{y}^{\text{test}} - \hat{\mathbf{y}}_r^{\text{test}}\right\|_2 \\ &= \arg\min_{r} \left\|\mathbf{y}^{\text{test}} - \mathbf{D}\Phi_r\left(\hat{\boldsymbol{\alpha}}^{\text{test}}\right)\right\|_2 \end{aligned} \tag{4}$$

where $\Phi_r(\hat{\boldsymbol{\alpha}}^{\text{test}})$ is an vector operator that preserves coefficients of $\hat{\boldsymbol{\alpha}}^{\text{test}}$ corresponding to the class r and sets all other coefficients to zero.

In order to further exploit the spatial context of the HSI, for each test pixel $\mathbf{y}^{\text{test}}$, [27]–[29] introduce a region-based SRC. Specifically, for each test pixel, the region-based SRC defines a local region of fixed size and shape. Then, pixels within the region are sparsely represented, and the corresponding sparse coefficients are used to determine the class label of the test pixel $\mathbf{y}^{\text{test}}$. Compared with the aforementioned pixelwise-based SRC, the region-based SRC can provide a better classification performance in terms of accuracies. However, the fixed-size region may not effectively exploit the spatial information of the HSI. For example, if the shape and size of the region are not appropriately selected, many pixels in the region may be uncorrelated to the test pixel, consequently deteriorating the classification accuracy. In addition, to classify each pixel, its multiple neighboring pixels within the local region should be coded. That will create a high computational burden. To address these issues, the proposed superpixel-based sparse representation model is introduced in the following section.

B. Discriminative K-SVD Dictionary Learning Algorithm

The performance of the sparse representation relies heavily on the quality of the dictionary $\mathbf{D}$. In [27]–[29], the dictionary is constructed by directly extracting pixels from the image of interest. In this way, the dictionary usually requires a large number of training pixels while still not possibly being robust to represent each pixel to be classified. Recently, many dictionary learning algorithms [44] have been proposed to achieve a compact representation. One widely used dictionary learning algorithm is the K-SVD [46] which aims to solve the following optimization problem:

$$\begin{aligned}\{\hat{\mathbf{D}}, \hat{\mathbf{A}}^{\text{train}}\} = \arg\min_{\hat{\mathbf{D}}, \hat{\mathbf{A}}^{\text{train}}} \|\mathbf{Y}^{\text{train}} - \mathbf{D}\mathbf{A}^{\text{train}}\|_F^2, \\ \text{subject to } \|\boldsymbol{\alpha}_z^{\text{train}}\|_0 \leqslant K, z = 1, \ldots, Z\end{aligned} \tag{5}$$

where $\mathbf{Y}^{\text{train}}$ is composed of the training samples $[\mathbf{y}_1^{\text{train}}, \ldots, \mathbf{y}_Z^{\text{train}}]$, $\boldsymbol{\alpha}_z^{\text{train}}$ is the sparse coefficient for the training sample $\mathbf{y}_z^{\text{train}}$, $\mathbf{A}^{\text{train}}$ is the corresponding sparse coefficients matrix $[\boldsymbol{\alpha}_1^{\text{train}}, \ldots, \boldsymbol{\alpha}_Z^{\text{train}}]$, Z is the number of training samples, and $\|\mathbf{Y}^{\text{train}} - \mathbf{D}\mathbf{A}^{\text{train}}\|_F^2$ is the Frobenius norm of $(\mathbf{Y}^{\text{train}} - \mathbf{D}\mathbf{A}^{\text{train}})$. The K-SVD splits the problem in (5) into sparse coding and dictionary updating stages, which are solved within an iterative loop. In the sparse coding, $\mathbf{D}$ is initially kept fixed, and the sparse coefficients matrix $\mathbf{A}^{\text{train}}$ is computed via the OMP algorithm [47]. Then, given the sparse coefficients matrix, $\mathbf{A}^{\text{train}}$, the K-SVD updates one atom at a time via SVD [46]. The K-SVD improves the representative power of the dictionary and has been demonstrated to work well in many image restoration problems [46], [52], [53].

Recently, by utilizing the class information of training samples, a discriminative K-SVD dictionary learning algorithm was proposed for classification [45]. Such an algorithm introduces a classification error constraint $\|\mathbf{H}^{\text{train}} - \mathbf{W}\mathbf{A}^{\text{train}}\|_F^2$, where $\mathbf{H}^{\text{train}} = [\mathbf{h}_1^{\text{train}}, \ldots, \mathbf{h}_Z^{\text{train}}] \in \mathbb{R}^{R\times Z}$ is the class matrix of the training samples $\mathbf{Y}^{\text{train}}$ and $\mathbf{W} \in \mathbb{R}^{R\times N}$ is a classifier which can obtain the class vector $\mathbf{h}_z^{\text{train}}$ of $\mathbf{y}_z^{\text{train}}$ with the sparse coefficient $\boldsymbol{\alpha}_z^{\text{train}}$. Furthermore, $\mathbf{h}_z^{\text{train}} = [0, 0, \ldots, 1, \ldots, 0, 0]^T \in \mathbb{R}^R$ is a class vector corresponding to one training sample $\mathbf{y}_z^{\text{train}}$ and the nonzero position in $\mathbf{h}_z^{\text{train}}$ represents the class label r_z^{train} of $\mathbf{y}_z^{\text{train}}$. Then, the classification error constraint $\|\mathbf{H}^{\text{train}} - \mathbf{W}\mathbf{A}^{\text{train}}\|_F^2$ can be incorporated into (5) to jointly learn the discriminative dictionary $\mathbf{D}$ and the classifier $\mathbf{W}$ [45]

$$\begin{aligned}\{\hat{\mathbf{D}}, \hat{\mathbf{A}}^{\text{train}}, \hat{\mathbf{W}}\} = \arg\min_{\hat{\mathbf{D}}, \hat{\mathbf{A}}^{\text{train}}, \hat{\mathbf{W}}} \|\mathbf{Y}^{\text{train}} - \mathbf{D}\mathbf{A}^{\text{train}}\|_F^2 \\ + \lambda_1 \left\|\mathbf{H} - \mathbf{W}\mathbf{A}^{\text{train}}\right\|_F^2, \\ \text{subject to } \|\boldsymbol{\alpha}_z^{\text{train}}\|_0 \leqslant K, z = 1, \ldots, Z\end{aligned} \tag{6}$$

where λ_1 is a scalar balancing the relative contribution of the reconstruction and classification error constraints. As described in [45], (6) can be rewritten as

$$\begin{aligned}&\{\hat{\mathbf{D}}, \hat{\mathbf{A}}^{\text{train}}, \hat{\mathbf{W}}\} \\ &= \arg\min_{\hat{\mathbf{D}}, \hat{\mathbf{A}}^{\text{train}}, \hat{\mathbf{W}}} \left\| \begin{pmatrix} \mathbf{Y}^{\text{train}} \\ \sqrt{\lambda_1}\mathbf{H}^{\text{train}} \end{pmatrix} - \begin{pmatrix} \mathbf{D} \\ \sqrt{\lambda_1}\mathbf{W} \end{pmatrix} \mathbf{A}^{\text{train}} \right\|_2^2 \\ &\quad \text{subject to} \|\boldsymbol{\alpha}_z^{\text{train}}\|_0 \leqslant K, z = 1, \ldots, Z.\end{aligned} \tag{7}$$

Let us now assume that $\mathbf{Y}_* = (\mathbf{Y}^{\text{train}}/\sqrt{\lambda_1}\mathbf{H}^{\text{train}})$ and $\mathbf{D}_* = (\mathbf{D}/\sqrt{\lambda_1}\mathbf{W})$. Then, (7) is equivalent to the following optimization problem:

$$\begin{aligned}\{\hat{\mathbf{D}}_*, \hat{\mathbf{A}}^{\text{train}}\} = \arg\min_{\hat{\mathbf{D}}_*, \hat{\mathbf{A}}} \|\mathbf{Y}_* - \mathbf{D}_*\mathbf{A}^{\text{train}}\|_F^2 \\ \text{subject to} \|\boldsymbol{\alpha}_z^{\text{train}}\|_0 \leqslant K, z = 1, \ldots, Z\end{aligned} \tag{8}$$

which can be also solved by the K-SVD algorithm.

III. SBDSM for HSI Classification

A superpixel can be regarded as a small spatial region, whose shape and size can be adaptively adjusted according to different spatial structures. Instead of classifying each single pixel as in [27]–[29], the proposed SBDSM algorithm is designed for the classification of the superpixel in HSI with a discriminative dictionary. Generally, the proposed SBDSM is composed of the following four parts, which are further described in the following sections: 1) superpixel map creation; 2) superpixel-based sparse representation; c) discriminative dictionary learning; and d) superpixel classification.

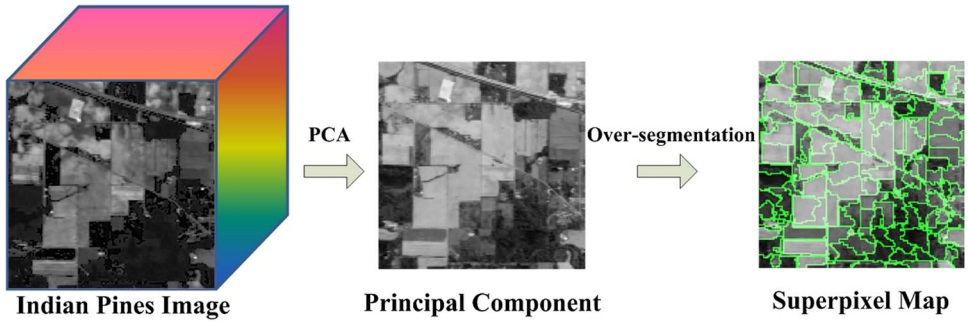


Fig. 1. Procedure for the superpixel creation.

A. Superpixel Map Creation

The superpixel map is created by applying an efficient oversegmentation approach [40] on the HSI. Then, the HSI can be clustered into L nonoverlapping superpixels. To reduce the computational cost, before the segmentation, principal component analysis (PCA) [54] is applied on the original HSI. Since the principal component corresponding to the highest eigenvalue (i.e., the first principal component) should contain the most important information in terms of variation for the whole HSI, it is used as the base image for the oversegmentation. The procedure for creating the superpixel map is illustrated in Fig. 1.

B. Superpixel-Based Sparse Representation

When the superpixel map has been created, it can be utilized with the original HSI to extract superpixels. Each superpixel is composed of a number of spectral pixels $[\mathbf{y}_{i,1}, \mathbf{y}_{i,2}, \ldots]$, which can be arranged into a matrix $\mathbf{Y}_i^{\mathrm{SP}}$. Then, pixels within each superpixel are assumed to have very similar spectral characteristics and their correlations can be exploited by joint sparse regularization [41]. Specifically, the joint regularization requires that pixels within each superpixel $\mathbf{Y}_i^{\mathrm{SP}}$ shall be simultaneously decomposed by linear combinations of a few common atoms from a dictionary. Let the sparse coefficients of the $\mathbf{Y}_i^{\mathrm{SP}}$ be a matrix $\mathbf{A}_i^{\mathrm{SP}}$. The joint sparse regularization places a $\ell_{\mathrm{row},0}$-norm on the sparse matrix $\mathbf{A}_i^{\mathrm{SP}}$, as $\|\mathbf{A}_i^{\mathrm{SP}}\|_{\mathrm{row},0}$. The $\ell_{\mathrm{row},0}$-norm denotes the sparse norm, which is used to select a number of the most representative nonzero rows in $\mathbf{A}_i^{\mathrm{SP}}$. We should note that the coefficient values in each row vector may be varied to cover the slight differences in similar pixels, as illustrated in Fig. 2(b). By incorporating the joint sparse regularization into (3), the SBDSM pursues the joint sparse matrix $\mathbf{A}_i^{\mathrm{SP}}$ by solving the following optimization problem:

$$\hat{\mathbf{A}}_i^{\mathrm{SP}} = \arg\min \left\|\mathbf{Y}_i^{\mathrm{SP}} - \mathbf{D}\mathbf{A}_i^{\mathrm{SP}}\right\|_2 \text{ subject to } \left\|\mathbf{A}_i^{\mathrm{SP}}\right\|_{\mathrm{row},0} \leqslant K. \tag{9}$$

In this paper, a variant of the OMP algorithm, called the simultaneous OMP (SOMP) [41], is used to efficiently solve the aforementioned problem. A schematic of the superpixel-based sparse representation is shown in Fig. 3.

Note that, enforcing the structured constraints (e.g., manifold [55] and mixed $l_{1,2}$-norm [56]) on (9) may produce a better sparse coefficients matrix, but also create higher computational cost.

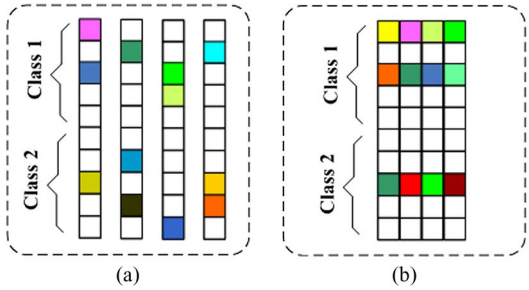


Fig. 2. Illustration of two different sparsity patterns in a sparse coefficients matrix $\mathbf{A}_i^{\mathrm{SP}}$. Each column represents one sparse coefficient vector and each square block denotes a coefficient value. The white blocks denote zero values, whereas the color blocks represent nonzero values. (a) Separate sparsity patterns in $\mathbf{A}_i^{\mathrm{SP}}$. (b) Joint sparsity patterns in $\mathbf{A}_i^{\mathrm{SP}}$ with the joint sparse regularization. The joint sparse regularization makes the nonzero coefficients of the matrix $\mathbf{A}_i^{\mathrm{SP}}$ belong to the same row although their values may be varied.

C. Discriminative Dictionary Learning With Class-Labeled OMP Algorithm

To achieve a discriminative representation, the learning algorithm in (6) can be directly utilized to train one dictionary $\mathbf{D}$, as well as one classifier $\mathbf{W}$. As previously described, (6) can be optimized by the K-SVD, but with a significant computational burden. The main computational cost of the K-SVD is caused by the OMP, which needs to search through all the dictionary atoms in every iteration to pursue the sparse coefficients [48]. In addition, the sparse coefficients obtained by the OMP algorithm lack the discriminability for training a simple linear classifier $\mathbf{W}$ [49], [50]. By utilizing the class information of the training samples and dictionary atoms, Jiang *et al.* [49], [50] proposed a novel discriminative sparse coefficients constraint $\|\mathbf{Q} - \mathbf{P}\mathbf{A}^{\mathrm{train}}\|_F^2$, where $\mathbf{Q} \in \mathbb{R}^{N \times Z}$ is the discriminative sparse coefficients matrix with both training pixels and dictionary atom labels, and $\mathbf{P}$ is a transformation matrix, which transforms the original sparse coefficients $\mathbf{A}^{\mathrm{train}}$ to the discriminative counterparts $\mathbf{Q}$. Such a constraint can then be incorporated into (6) to ensure that training samples from the same class have similar sparse coefficients, which produces a well-trained classifier. However, when $\mathbf{Q}$ is further concatenated with $\mathbf{W}$ and $\mathbf{D}$ in a joint optimization process, the dimensionality will significantly increase and, thus, a very high computational burden is created.

To address the aforementioned issues, inspired by works in [49], [50], a class-labeled OMP algorithm is proposed here for solving (8) by incorporating the class information of the

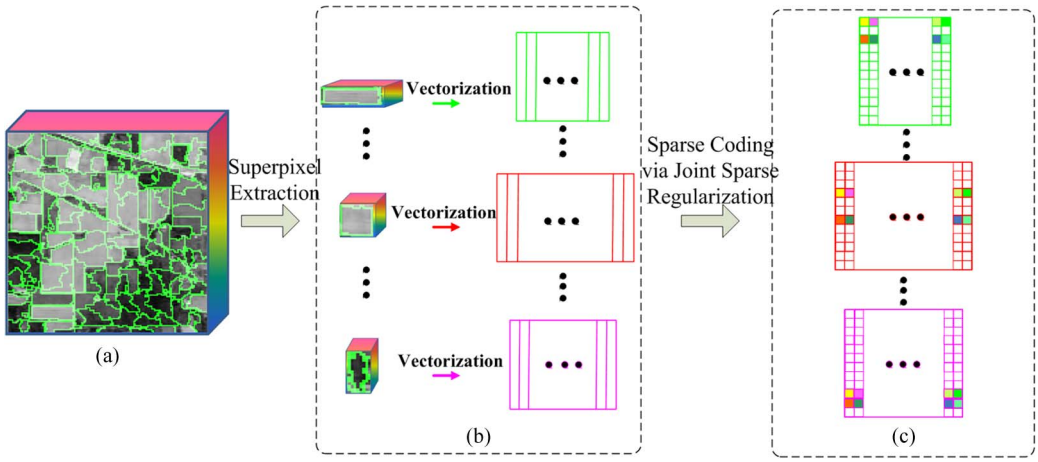


Fig. 3. Schematic of the superpixel-based sparse representation.

Class Labeled OMP Algorithm

Input: $\mathbf{y}_z^{train}$, one training sample; r_z^{train}, class label for the training sample; $\mathbf{D}$, initial dictionary; $\mathbf{Q}$, class labels for the dictionary atoms; K, sparsity level.

Output: $\boldsymbol{\alpha}_z^{train}$, sparse coefficient for the $\mathbf{y}_z^{train}$.

Initialization: set the index set $\mathbf{I}^1 = \varnothing$, residual $\mathbf{u}^1 = \mathbf{y}_z^{train}$, the iteration counter $iter = 1$.

1: Find the index set $\mathbf{J}$ of atoms with the same class labels for the training sample: $\mathbf{J} = \text{find}\left(\mathbf{Q} == r_z^{train}\right)$.

2: Compute residual correlation vector $\mathbf{E}^{iter}$: $\mathbf{E}^{iter} = \mathbf{D}_\mathbf{J}^T \mathbf{u}^{iter}$, where $\mathbf{D}_\mathbf{J}$ denotes the subdictionary constructed by atoms from the set $\mathbf{J}$.

3: Find the index of the best representative atom (in the subdictionary $\mathbf{D}_\mathbf{J}$): $\hat{j}^{iter} = \max_j \left|\mathbf{E}_j^{iter}\right|, j \in \mathbf{J}$.

3: Merge the index of the newly selected atom with the previously selected index set: $\mathbf{I}^{iter} = \mathbf{I}^{iter} \bigcup \hat{j}^{iter}$.

4. Estimate the sparse coefficient $\boldsymbol{\alpha}_{z,iter}^{train}$: $\boldsymbol{\alpha}_{z,iter}^{train} = \left(\mathbf{D}_{\mathbf{I}^{iter}}^T \mathbf{D}_{\mathbf{I}^{iter}}\right)^{-1} \left(\mathbf{D}_{\mathbf{I}^{iter}}\right)^T \mathbf{y}_z^{train}$.

5: Update the residual: $\mathbf{u}^{iter} = \mathbf{y}_z^{train} - \mathbf{D}\boldsymbol{\alpha}_{z,iter}^{train}$.

If iteration counter $iter > K$, stop the procedures and output the final sparse coefficient $\boldsymbol{\alpha}_{z,iter}^{train}$; otherwise, set $iter = iter+1$ and go to the step 2.

Fig. 4. Class-labeled OMP algorithm.

training samples and the dictionary atoms into the original OMP algorithm. The class-labeled OMP is based on the assumption that each training sample should be ideally represented by dictionary atoms from the same class. To pursue the sparse coefficient for each training sample, the class-labeled OMP only searches a small part of the atoms associated with the same class in the training samples and thus greatly reduces the computational cost. Accordingly, the training samples from a particular class will be represented by atoms from the same class and, therefore, their sparse coefficients are forced to be similar. This way, the discriminability of the sparse coefficients can be also enhanced to train the classifier in an appropriate manner. The whole class-labeled OMP algorithm is detailed in Fig. 4. After obtaining the sparse coefficients for the training samples, the dictionary and the classifier can be updated by SVD decomposition.

D. Classification With the Learned Dictionary and Classifier

After $\mathbf{D}$ and $\mathbf{W}$ are obtained by solving (6), they cannot be directly used for classification, since they are jointly normalized in $\mathbf{D}_*$ in the aforementioned optimization algorithm. Following [45], the desired dictionary $\hat{\mathbf{D}}$ and classifier $\hat{\mathbf{W}}$ should be computed as

$$\hat{\mathbf{D}} = \left\{ \frac{\mathbf{d}_1}{\|\mathbf{d}_1\|_2}, \frac{\mathbf{d}_2}{\|\mathbf{d}_2\|_2}, \ldots, \frac{\mathbf{d}_N}{\|\mathbf{d}_N\|_2} \right\}$$
$$\hat{\mathbf{W}} = \left\{ \frac{\mathbf{w}_1}{\|\mathbf{w}_1\|_2}, \frac{\mathbf{w}_2}{\|\mathbf{w}_2\|_2}, \ldots, \frac{\mathbf{w}_N}{\|\mathbf{w}_N\|_2} \right\} \tag{10}$$

where $\mathbf{d}_n$ and $\mathbf{w}_n$ are the nth column of $\mathbf{D}$ and $\mathbf{W}$, respectively.

For each superpixel $\mathbf{Y}_i^{\text{SP}} = [\mathbf{y}_{i,1}, \mathbf{y}_{i,2}, \ldots]$, the corresponding joint sparse coefficient matrix $\mathbf{A}_i^{\text{SP}} = [\boldsymbol{\alpha}_{i,1}, \boldsymbol{\alpha}_{i,2}, \ldots]$ is first computed by addressing the problem in (9). Then, the classifier $\hat{\mathbf{W}}$ can be applied on $\mathbf{A}_i^{\text{SP}}$ to create a class matrix $\mathbf{H}_i^{\text{SP}} = [\mathbf{h}_{i,1}, \mathbf{h}_{i,2}, \ldots]$ for each test superpixel

$$\mathbf{H}_i^{\text{SP}} = \hat{\mathbf{W}} \times \mathbf{A}_i^{\text{SP}} \tag{11}$$

where $\mathbf{h}_{i,1} \in \mathbb{R}^R$ is the class vector of the corresponding pixel $\mathbf{y}_{i,1}$ in $\mathbf{Y}_i^{\text{SP}}$; and the position of the max value in $\mathbf{h}_{i,1}$ is the class label of the $\mathbf{y}_{i,1}$, as described in [45]. Finally, each row of the $\mathbf{H}_i^{\text{SP}}$ is summed together as a new class vector $\mathbf{h}_i^{\text{SP}} \in \mathbb{R}^R$, and the class label $\hat{r}_i^{\text{SP}}$ of each superpixel $\mathbf{Y}_i^{\text{SP}}$ can be determined by the position of the max value in $\mathbf{h}_i^{\text{SP}}$

$$\hat{r}_i^{\text{SP}} = \arg\max_r \left(\mathbf{h}_i^{\text{SP}}(r)\right), r = 1, \ldots, R. \tag{12}$$

The whole SBDSM algorithm is summarized in Fig. 5.

IV. Experimental Results

Here, the performance of the proposed algorithm is tested in classification of four hyperspectral datasets,[1] i.e., the AVIRIS Indian Pines image, AVIRIS Salinas image, ROSIS-03 University of Pavia image, and HYDICE Washington DC image. The proposed classification algorithm is split into four versions: SBDSM-NoDL, SBDSM-KSVD, SBDSM-Cseg, and SBDSM. In the superpixel creation step, the SBDSM-NoDL,

[1]Datasets can be downloaded at: http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes.

SBDSM Algorithm

Input: HSI image

1: Discriminative Dictionary and classifier Learning with Class Labeled OMP Algorithm:

Utilize the proposed class labeled OMP algorithm to solve (6) and obtain a discriminative dictionary $\hat{\mathbf{D}}$ and classifier $\hat{\mathbf{W}}$

2: Superpixel Map Creation:

Apply the PCA on the HSI test image to reduce the dimensionality and then use the over-segmentation approach [40] to create a superpixel map

3: Superpixel Based Sparse Representation:

For each superpixel $\mathbf{Y}_i^{\text{superpixel}}$ of the HSI, use the SOMP algorithm [41] to solve (9) with the trained dictionary $\hat{\mathbf{D}}$ and obtain the corresponding sparse matrix $\hat{\mathbf{A}}_i^{\text{superpixel}}$

4: Classification of each superpixel with the Learned Classifier:

Apply the classifier $\hat{\mathbf{W}}$ on the sparse matrix $\hat{\mathbf{A}}_i^{\text{superpixel}}$ to obtain the class label of the corresponding superpixel $\mathbf{Y}_i^{\text{superpixel}}$, as in (11) and (12). Use the class labels of all the superpixels to constitute the final classification map

Output: Classification map

Fig. 5. Summary of the SBDSM algorithm.

SBDSM-KSVD, and SBDSM methods utilize a recent oversegmentation algorithm [40], whereas the SBDSM-Cseg method adopts a conventional oversegmentation algorithm [57]. To train the discriminative dictionary, the SBDSM and SBDSM-Cseg adopted the dictionary-learning method with the proposed class-labeled OMP algorithm, whereas the SBDSM-KSVD utilized the K-SVD learning method in [45]. By contrast, the SBDSM-NoDL directly extracted pixels from HSI as the sparse dictionary. In the sparse representation stage, the aforementioned four methods all solve the optimization problem in (9). For the classification of each superpixel $\mathbf{Y}_i^{\text{SP}}$, the SBDSM-KSVD, SBDSM-Cseg, and SBDSM address the problems in (11) and (12), whereas the SBDSM-NoDL solves the following problem:

$$\hat{r}_i^{\text{SP}} = \arg\min_r \left\| \mathbf{Y}_i^{\text{SP}} - \mathbf{D}\Phi_r\left(\hat{\mathbf{A}}_i^{\text{test}}\right) \right\|_2 \tag{13}$$

where $\Phi_r(\hat{\mathbf{A}}_i^{\text{test}})$ is a matrix operator that preserves coefficients of $\hat{\mathbf{A}}_i^{\text{test}}$, corresponding to the class r and sets all other coefficients to zero. In addition, the experiments included other seven competitive algorithms for comparison: SVM [2], extended morphological profile (EMP) [14], Logistic regression via splitting and augmented Lagrangian-multilevel logistic (LORSAL-MLL) [58], Pixelwise SRC [27], Region-based SRC [27], Pixel-LCKSVD [50], and Superpixel-SVM. The Superpixel-SVM is a combination of the superpixel segmentation and SVM classifier, which first uses the method in Section II-A to create the superpixels and then applies the SVM to determine the class label of each superpixel. Since the original SVM can only classify each spectral pixel, majority voting is used to fuse the SVM's results within each superpixel. The SVM classifier was implemented with the LIBSVM library [59] and the spectral-only Gaussian kernel was adopted without considering spatial information. For EMP and LORSAL-MLL, the local spatial context of HSI was exploited by the use of the extended morphological profile and the multilevel logistic prior-based segmentation technique, respectively. The pixelwise SRC, region-based SRC, and Pixel-LCKSVD are three sparsity-based classifiers that aim to classify one spectral pixel at a time. The Pixelwise SRC and Pixel-LCKSVD only exploits the spectral information, whereas the Region-based SRC utilizes both the spectral information and the spatial context of the fixed-size local region centered at each test pixel. The Pixel-LCKSVD adopted the LCKSVD algorithm [50] for training of the dictionary while the Pixelwise SRC and Region-based SRC directly extracted pixels from HSI for the sparse dictionary.

A. Data Set Description

The Indian Pines image was captured by Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over the agricultural Indian Pines test site of northwestern Indiana. The image is of size $145 \times 145 \times 220$ with a spatial resolution of 20 m per pixel and a spectral coverage ranging from 0.2 to 2.4 μm. In the experiments, the number of spectral bands was reduced to 200 by removing 20 water absorption bands [60]. The image contains 16 reference classes, most of which are different types of crops (e.g., corns, soybeans, and wheat). Fig. 6(a) and (b) show the color composite of the Indian Pines image and the corresponding reference data.

The Salinas image, which captures the area of the Salinas Valley, California, was also acquired by the AVIRIS sensor. The image has a spatial resolution of 3.7 m per pixel and is of size $512 \times 217 \times 224$. As for the Indian Pines image, 20 water absorption spectral bands are discarded and the reference image contains 16 different classes. Fig. 7(a) and (b) show the color composite of the Salinas image and the corresponding reference data.

The University of Pavia image was recorded by the Reflective Optics System Imaging Spectrometer (ROSIS-03) sensor over an urban area surrounding the University of Pavia, Italy. The ROSIS-03 sensor in this case generated an image with a spatial resolution of 1.3 m per pixel and a spectral coverage ranging from 0.43 to 0.86 μm [61]. The University of Pavia image consists of 610×340 pixels, each including 103 spectral bands with the nosiest bands discarded. Nine reference classes are considered for this image. Fig. 8(a) and (b) show the color composite of the University of Pavia image and the corresponding reference data.

The Washington DC image was captured by the Hyperspectral Digital Image Collection Experiment (HYDICE) sensor over the Washington DC Mall. The image is of size $280 \times 307 \times 210$ with a spectral coverage ranging from 0.2 to 2.4 μm. In the experiments, the number of spectral bands was reduced to 191 by removing bands ranging from 0.9 to 1.4 μm, since the atmosphere of these bands is opaque. The reference of this image contains six classes. Fig. 9(a) and (b) show the color composite of the Washington DC image and the corresponding reference data.

B. Quantitative Metrics

The OA, average accuracy (AA), and Kappa coefficient are adopted as the objective metrics to evaluate the classification results. The OA is computed by the percentage of correctly classified test pixels. The AA is mean of the percentage of correctly classified pixels for each class. The Kappa coefficient shows the percentage of classified pixels corrected by the number of agreements that would be expected purely by chance.

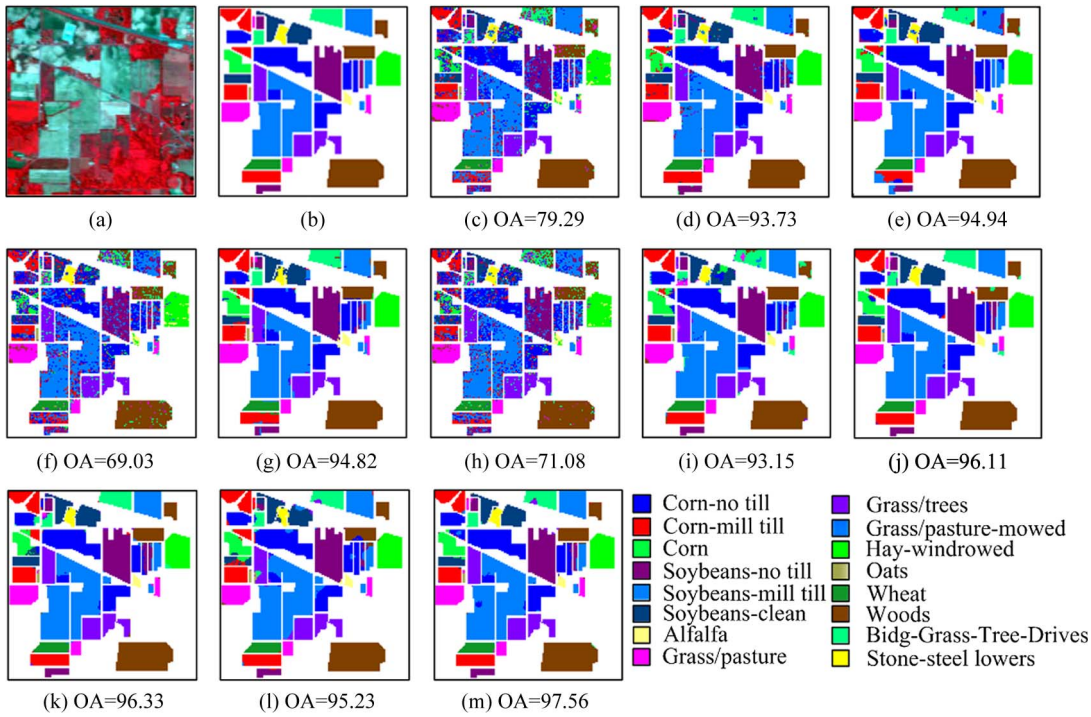


Fig. 6. Indian Pines image. (a) Three-band color composite image. (b) Reference image, and the classification results (OA in percentage) obtained by the (c) SVM [2], (d) EMP [14], (e) LORSAL-MLL [58], (f) Pixelwise SRC [27], (g) Region-based SRC [27], (h) Pixel-LCKSVD [50], (i) Superpixel-SVM, (j) SBDSM-NoDL, (k) SBDSM-KSVD [45], (l) SBDSM-Cseg [57], and (m) SBDSM methods.

In addition, the F-Measure [62] is used as the quantitative metric to evaluate the oversegmentation results for creating the superpixel. The F-Measure is computed by $2RP/(R+P)$, where R and P are recall and precision of the segmentation result relative to the ground reference map, respectively. Specifically, if A_{SR} is the set of pixels from the segmentation result and $A_{\mathrm{GR}}$is the set of pixels from the ground reference map, $R = |A_{\mathrm{SR}} \cap A_{\mathrm{GR}}|/A_{\mathrm{GR}}$ and $P = |A_{\mathrm{SR}} \cap A_{\mathrm{GR}}|/A_{\mathrm{SR}}$.

C. Comparison of Results

In the experiments, the parameters are empirically selected for different versions of the proposed SBDSM algorithm, i.e., SBDSM-NoDL, SBDSM-KSVD, SBDSM-Cseg, and SBDSM. For the four considered images, the sparsity level K in (9) is chosen to be 3. The number of superpixels L is selected as 600, 300, 1000, 1300 for Indian Pines, Salinas, University of Pavia, and Washington DC, respectively. The dictionary size per class N_r is selected as 80% of the number of the training samples in each class. In the following section, we will analyze the effects of the sparsity level, the dictionary size per class, and the number of superpixels on the performance of the proposed SBDSM method. Parameters C and $\sigma$for the SVM are obtained by fivefold cross validation. Specifically, parameter C was determined as 100, 10, 100, 1000 for the Indian Pines, Salinas, University of Pavia, and Washington DC images, respectively. The σ of the Gaussian kernel in the SVM was determined as 0.125, 0.25, 0.125, 0.125, on the Indian Pines, Salinas, University of Pavia, and Washington DC images, respectively. The parameters for the EMP and LORSAL-MLL are set to the default values as in [14] and [58] and the parameters for Pixelwise SRC and Region-based SRC are tuned to reach the best results in these experiments. For the Superpixel-SVM, the number of superpixels is set to be the same as in the proposed SBDSM algorithm, whereas parameters C and σ for the SVM classifier are set to be the same as for the SVM classifier.

The first experiment was conducted on the Indian Pines image. For each class of this image, 10% of the labeled samples were randomly selected as the training set and the remaining 90% used as test samples (see the third and fourth columns of Table I). The classification maps obtained from various techniques on the Indian Pines image are illustrated in Fig. 6. It can be observed that the SVM, Pixelwise SRC, and Pixel-LCKSVD that only consider the spectral information present a very noisy estimation in their classification maps. By incorporating the local spatial context of the HSI, the EMP, and Region-based SRC yield a smoother visual effect. However, these two approaches still fail to identify the pixels in the detailed and near-edge regions. By contrast, the proposed SBDSM approaches, i.e., SBDSM-NoDL, SBDSM-KSVD, and SBDSM, can provide a smooth appearance while still achieving higher classification accuracies in the detailed regions. The quantitative results for different methods are tabulated in Table II. Note that for each method, ten experiments with different randomly sampled training data were conducted and the mean and standard deviation of the classification accuracies are reported. As can be observed, the SBDSM-NoDL, SBDSM-KSVD, and SBDSM approaches outperform other compared methods in terms of the OA, AA, and the Kappa coefficient. In addition, it can be seen that the SBDSM that constructs the dictionary with the proposed discriminative learning algorithm performs better in terms of accuracies than both the SBDSM-KSVD that trains

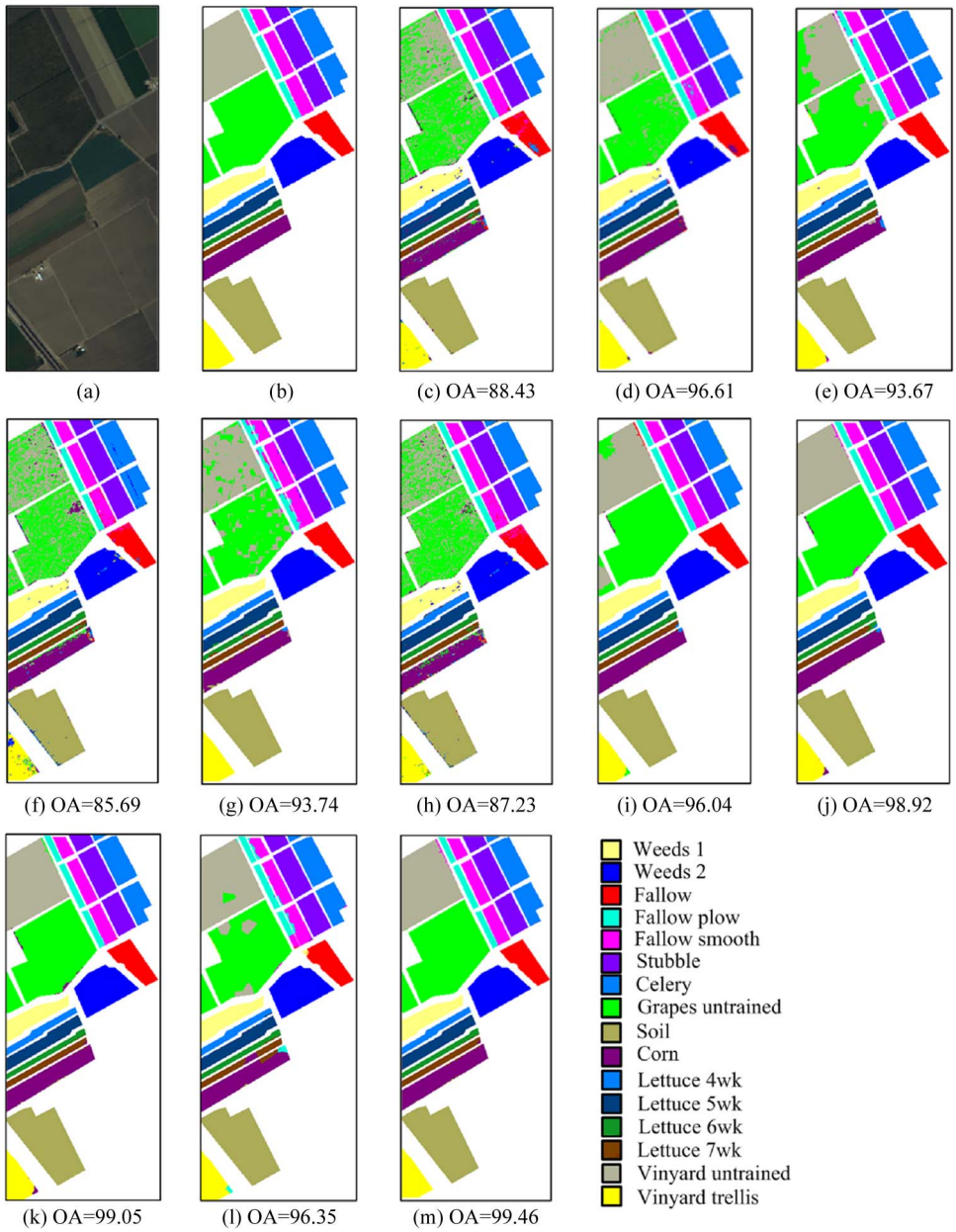


Fig. 7. Salinas Image. (a) Three-band color composite image. (b) Reference image, and the classification results (OA in percentage) obtained by the (c) SVM [2], (d) EMP [14], (e) LORSAL-MLL [58], (f) Pixelwise SRC [27], (g) Region-based SRC [27], (h) Pixel-LCKSVD [50], (i) Superpixel-SVM, (i) SBDSM-NoDL, (k) SBDSM-KSVD [45], (l) SBDSM-Cseg [57], and (m) SBDSM methods.

the dictionary with the original K-SVD algorithm [45] and the SBDSM-NoDL that constructs the dictionary with directly extracted pixels. This demonstrates, that compared with the K-SVD, the proposed learning algorithm can enable the dictionary and the sparse coefficients to be more discriminable for HSI classification.

The second and third experiments were performed on the Salinas and University of Pavia images, respectively. In the Salinas image, 1% of the labeled data were randomly chosen as the training samples and the rest 99% of data as the test set (see the sixth and seventh columns of Table I). As in some recent papers [31], [63], [64], that classified the University of Pavia image, we adopted a fixed number of training samples for each class. Specifically, in the experiments performed here, we randomly selected 250 spectral pixels from each class for training and used the remaining pixels as the test set (see the ninth and tenth columns of Table I). The training samples constitute only about 5% of the whole labeled data, which provides a comparatively challenging test set. The classification maps and quantitative results (mean and standard deviation over ten experiments) from different methods on the Salinas and University of Pavia images are shown in Figs. 7 and 8, and Tables III and IV. As shown from Fig. 7 and Table III, the proposed SBDSM approach and two approaches based on it, i.e., the SBDSM-KSVD and SBDSM-NoDL methods, perform better than the other classifiers on the Salinas image in terms of the visual quality of the classification map, as well as quantitative metrics on the Salinas image. In addition, we can observe from Fig. 8 and Table IV that the SBDSM-KSVD, SBDSM-NoDL, EMP, and LORSAL-MLL methods are generally superior to the

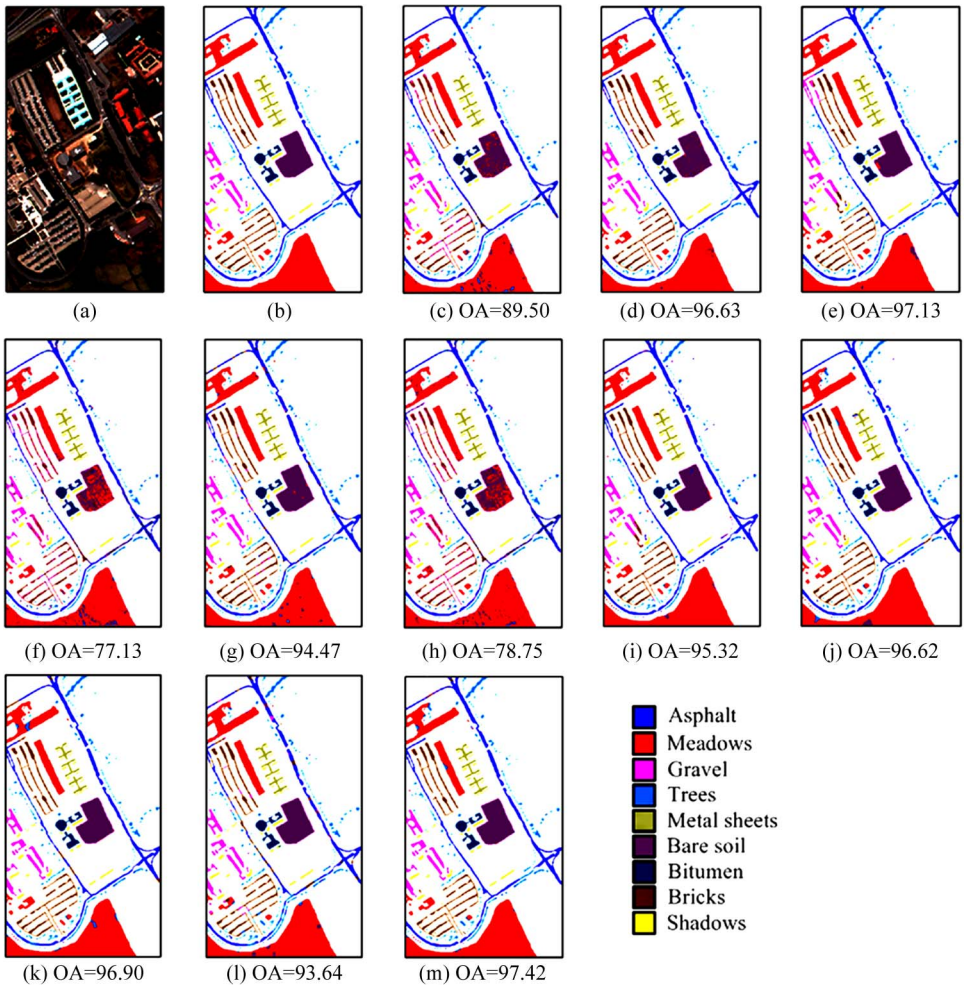


Fig. 8. Indian Pines image. (a) Three-band color composite image. (b) Reference image, and the classification results (OA in percentage) obtained by the (c) SVM [2], (d) EMP [14], (e) LORSAL-MLL [58], (f) Pixelwise SRC [27], (g) Region-based SRC [27], (h) Pixel-LCKSVD [50], (i) Superpixel-SVM, (i) SBDSM-NoDL, (k) SBDSM-KSVD [45], (l) SBDSM-Cseg [57], and (m) SBDSM methods.

SVM, Pixelwise SRC, Region-based SRC, and the Superpixel-SVM classifiers on the University of Pavia image. Moreover, utilizing both the superpixel-based sparse representation and the discriminative learning algorithm, the proposed SBDSM algorithm delivers higher classification accuracies compared with the SBDSM-KSVD, SBDSM-NoDL, EMP, and LORSAL-MLL methods on the University of Pavia image.

The fourth experiment was conducted on the Washington DC image. For this image, about 4% of the labeled data were randomly chosen for training, and the other 96% were used as a test set (see the 12th and 13th columns of Table I). The qualitative and quantitative results (mean and standard deviation over ten experiments) from different methods in applied on the Washington DC images are shown in Fig. 9 and Table V. As can be observed, the proposed SBDSM method can generally deliver better results than the other compared methods, in terms of visual quality and quantitative metrics. In addition, the SBDSM-Cseg with the conventional oversegmentation algorithm [57] does not provide very good results because the conventional oversegmentation method used in SBDSM-Cseg cannot create a very accurate superpixel map. Furthermore, unlike the aforementioned three test images, the superiority of the proposed superpixel-based SBDSM method over the Pixelwise-SRC and Pixel-LCKSVD methods on the Washington DC image is not very significant. This is mainly due to the reason that the Washington DC image has many detailed regions; therefore, there may not be enough spatial information to be exploited for enhancing the performance, as compared with the aforementioned three images.

D. Running Time Comparison

The average running times (over ten realizations) of the proposed SBDSM method and other compared methods for classifying the four-test HSI images are reported on Table VI. The execution time for the proposed SBDSM method reported in Table VI includes the time consumed in the superpixel map creation, dictionary learning, and sparse classification stages. All the programs are executed on a laptop computer with an Intel Corei7-3720 CPU 2.60 GHz and 16 GB of RAM. As can be observed, the SBDSM-NoDL, SBDSM-KSVD, and SBDSM algorithms require much less running time compared with other methods. This demonstrates the high efficiency of the superpixel-based sparse classification strategy. It should be noted that, since the conventional segmentation algorithm in [57] is computationally very costly, the SBDSM-Cseg method

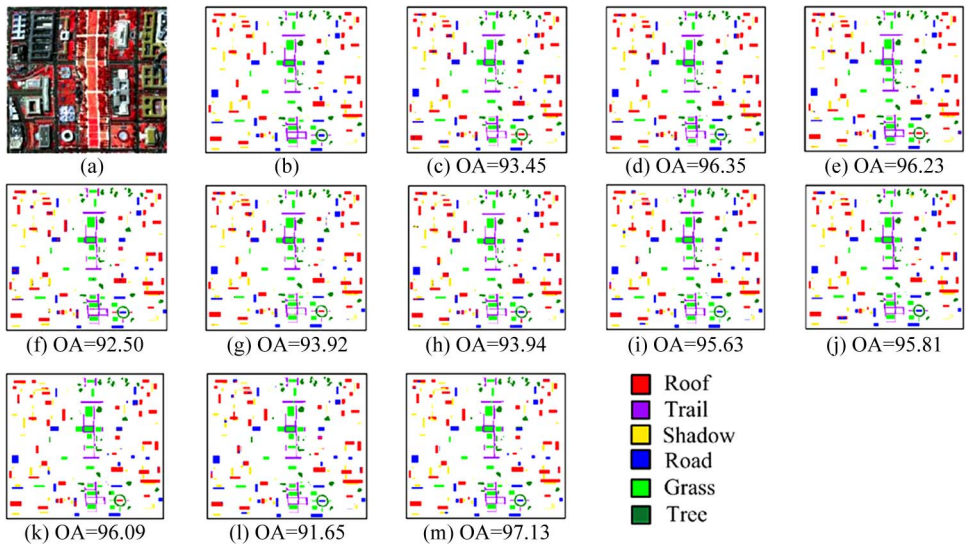


Fig. 9. Washington DC image. (a) Three-band color composite image. (b) Reference image, and the classification results [overall accuracy (OA) in percentage] obtained by the (c) SVM [2], (d) EMP [14], (e) LORSAL-MLL [58], (f) Pixelwise SRC [27], (g) Region-based SRC [27], (h) Pixel-LCKSVD [50], (i) Superpixel-SVM, (i) SBDSM-NoDL, (k) SBDSM-KSVD [45], (l) SBDSM-Cseg [57], and (m) SBDSM methods.

TABLE I
REFERENCE CLASSES FOR THE INDIAN PINES IMAGE, SALINAS IMAGE, UNIVERSITY OF PAVIA IMAGE, AND WASHINGTON DC IMAGE

Class	Indian Pines			Salinas			University of Pavia			Washington DC		
	Name	Train	Test	Name	Train	Test	Name	Train	Test	Name	Train	Test
1	Alfalfa	10	36	Weeds_1	20	1989	Asphalt	250	6381	Roof	125	3004
2	Corn-no till	143	1285	Weeds_2	37	3689	Meadows	250	18399	Grass	72	1718
3	Corn-min	83	747	Fallow	20	1956	Gravel	250	1849	Road	56	1346
4	Corn	24	213	Fallow plow	14	1380	Trees	250	2814	Trail	50	1214
5	Grass/Pasture	48	435	Fallow smooth	27	2651	Metal sheets	250	1095	Tree	48	1146
6	Grass/Trees	73	657	Stubble	40	3919	Bare soil	250	4779	Shadow	45	1075
7	Grass/Pasture-mowed	10	18	Celery	36	3543	Bitumen	250	1080	–	–	–
8	Hay-windrowed	48	430	Grapes	113	11158	Bricks	250	3432	–	–	–
9	Oats	10	10	Soil	62	6141	Shadows	250	697	–	–	–
10	Soybeans-no till	97	875	Corn	33	3245	–	–	–			
11	Soybeans-min	246	2209	Lettuce 4wk	11	1057	–	–	–			
12	Soybean-clean	59	534	Lettuce 5wk	19	1908	–	–	–			
13	Wheat	21	184	Lettuce 6wk	9	907	–	–	–	–	–	–
14	Woods	127	1138	Lettuce 7wk	11	1059	–	–	–	–	–	–
15	Building-Grass-Trees-Drives	39	347	Vinyard untrained	73	7195	–	–	–	–	–	–
16	Stone-steel Towers	10	83	Vinyard trellis	18	1789	–	–	–	–	–	–
	Total	1048	9201	Total	543	53586	Total	2250	40526	Total	396	9503

runs much slower than the SBDSM-NoDL, SBDSM-KSVD, and SBDSM approaches. By utilizing the discriminative dictionary learning, the SBDSM-KSVD and SBDSM are faster than the SBDSM-NoDL. This is due to the compact dictionary $\mathbf{D}$ for the sparse representation and efficient classifier $\mathbf{W}$ for the classification process. In addition, the SBDSM can provide a shorter running time than the SBDSM-KSVD, which demonstrates the efficiency of the proposed class-labeled OMP algorithm. Furthermore, we should note that the dictionary learning and sparse classification parts of the proposed SBDSM algorithm were coded in MATLAB, which was not optimized for speed. The processing time is expected to be further reduced by coding the SBDSM algorithm with C++ and adopting a general purpose graphics processing unit.

E. Effect of the Sparsity Level, Dictionary Size, and Number of Superpixels

Here, we examine the effect of the sparsity level K, sub-dictionary size per class N_r and the number of superpixels L

TABLE II
CLASSIFICATION ACCURACIES OF THE INDIAN PINES IMAGE OBTAINED BY THE SVM [2], EMP [14], LORSAL-MLL [58], PIXELWISE SRC [27], REGION-BASED SRC [27], PIXEL-LCKSVD [50], SUPERPIXEL-SVM, SBDSM-NoDL, SBDSM-KSVD [45], SBDSM-Cseg [57], AND SBDSM METHODS. CLASS-SPECIFIC ACCURACIES ARE IN PERCENTAGE

Class	SVM	EMP	LORSAL-MLL	Pixel-wise SRC	Region-based SRC	Pixel-LC KSVD	Superpixel-SVM	SBDSM-NoDL	SBDSM-KSVD	SBDSM-Cseg	SBDSM
Alfalfa	68.80	97.50	83.88	58.88	96.03	70.16	91.23	97.72	**98.61**	89.93	95.55
Corn-notill	71.26	92.18	92.12	54.35	94.47	57.82	89.24	**97.03**	94.49	94.70	95.72
Corn-min	73.91	88.47	89.05	51.46	92.35	55.63	84.87	94.90	94.18	93.39	**96.29**
Corn	62.28	79.24	**95.58**	35.83	92.55	40.79	84.03	90.56	86.34	89.49	88.63
Grass/Pasture	88.30	94.57	90.85	81.39	93.33	83.48	89.74	94.63	92.25	94.16	**95.21**
Grass/Trees	86.44	98.04	**99.72**	91.11	94.87	91.77	98.93	98.50	97.96	95.74	99.09
Grass/Pasture-mowed	88.07	61.24	92.22	89.62	88.88	90.38	**97.00**	85.35	92.12	86.11	83.33
Hay-windrowed	90.89	**100**	99.90	90.82	99.55	92.14	99.44	99.83	99.00	99.33	**100**
Oats	77.77	82.54	98.00	69.33	80.71	75.00	**100**	69.09	70.83	66.25	70.00
Soybeans-notill	74.42	92.57	91.86	66.79	91.93	69.15	92.38	93.50	95.66	91.91	**95.81**
Soybeans-min	78.79	92.58	95.89	71.10	96.36	73.65	95.88	96.04	98.30	96.65	**98.43**
Soybean-clean	69.31	88.76	**97.15**	41.56	90.61	45.48	93.47	96.56	95.91	89.40	95.46
Wheat	91.84	**100**	99.56	90.07	89.13	89.88	99.49	97.33	98.00	98.57	97.93
Woods	92.60	99.24	97.66	90.31	98.21	90.68	98.58	99.53	99.28	**99.81**	99.51
Building-Grass-Trees -Drives	68.84	**98.50**	93.14	33.90	94.23	39.14	71.33	93.01	93.67	94.66	97.02
Stone-steel Towers	**99.05**	99.13	82.41	87.95	81.23	91.34	89.15	92.88	90.16	74.09	89.63
OA (Mean in %)	79.53	93.56	94.73	68.55	94.66	71.08	92.81	96.30	96.32	95.09	**97.12**
OA (Standard Dev.)	0.63	0.45	0.37	0.63	0.21	0.71	1.13	0.78	0.83	0.33	0.41
AA (Mean in %)	80.01	91.54	**93.69**	69.03	92.15	72.28	92.17	93.53	93.55	90.89	93.61
AA (Standard Dev.)	2.16	1.36	0.85	1.34	1.74	1.72	0.96	0.89	1.13	2.13	0.85
Kappa (Mean)	0.77	0.93	0.94	0.64	0.94	0.67	0.92	0.96	0.96	0.94	**0.97**
Kappa (Standard Dev.)	0.007	0.005	0.003	0.007	0.003	0.007	0.012	0.009	0.010	0.004	0.005

TABLE III
CLASSIFICATION ACCURACIES OF THE SALINAS IMAGE OBTAINED BY THE SVM [2], EMP [14], LORSAL-MLL [58], PIXELWISE SRC [27], REGION-BASED SRC [27], PIXEL-LCKSVD [50], SUPERPIXEL-SVM, SBDSM-NoDL, SBDSM-KSVD [45], SBDSM-Cseg [57], AND SBDSM METHODS. CLASS-SPECIFIC ACCURACIES ARE IN PERCENTAGE

Class	SVM	EMP	LORSAL-MLL	Pixel-wise SRC	Region-based SRC	Pixel-LC KSVD	Superpixel-SVM	SBDSM-NoDL	SBDSM-KSVD	SBDSM-Cseg	SBDSM
Weeds_1	99.74	99.84	99.44	98.20	**100**	98.27	**100**	**100**	**100**	**100**	**100**
Weeds_2	99.01	99.76	99.95	98.09	**99.98**	98.15	99.81	98.64	**99.98**	99.90	99.97
Fallow	91.05	93.15	99.78	94.13	97.61	95.43	92.86	97.73	**100**	99.94	99.31
Fallow plow	97.04	98.49	98.34	98.71	83.24	98.48	99.92	96.11	**99.92**	63.76	**99.92**
Fallow smooth	98.07	99.16	98.78	91.85	97.10	94.08	99.39	**99.36**	99.29	92.64	99.30
Stubble	**99.98**	99.98	99.83	99.68	97.63	99.51	99.81	99.92	99.92	99.63	99.92
Celery	98.89	**99.92**	99.66	99.25	99.57	99.13	98.71	99.78	99.84	99.47	99.88
Grapes	75.96	92.96	90.76	71.09	88.61	74.04	96.50	99.01	**99.57**	96.97	99.43
Soil	98.87	99.25	99.97	97.26	99.97	97.98	99.98	**100**	99.99	99.87	**100**
Corn	88.86	93.37	94.15	85.57	96.11	87.00	96.09	97.47	97.62	90.86	**97.70**
Lettuce 4wk	91.77	98.80	95.34	93.54	97.37	**93.49**	**100**	**100**	**100**	95.59	**100**
Lettuce 5wk	95.75	96.53	99.99	99.74	95.52	**99.87**	**100**	**100**	**100**	96.54	**100**
Lettuce 6wk	94.78	98.01	97.83	96.83	95.08	96.72	97.86	**98.08**	97.96	95.80	97.94
Lettuce 7wk	96.47	**97.30**	95.95	92.76	94.64	92.10	95.99	94.28	92.33	80.73	92.48
Vinyard untrained	72.35	91.74	73.55	59.12	84.07	61.42	78.43	99.20	**97.27**	95.90	99.54
Vinyard trellis	98.64	98.30	**98.92**	95.41	99.33	95.70	94.36	97.52	98.60	99.21	98.69
OA (Mean in %)	89.33	96.23	93.75	85.89	93.96	87.12	95.43	98.98	99.11	96.19	**99.37**
OA (Standard Dev.)	0.59	0.25	0.42	0.36	0.41	0.34	0.74	0.39	0.40	0.45	0.23
AA (Mean in %)	93.58	97.29	96.39	91.95	95.36	92.58	96.86	98.57	98.89	94.17	**99.01**
AA (Standard Dev.)	0.44	0.29	0.24	0.10	0.24	0.40	0.61	0.66	0.42	0.68	0.41
Kappa (Mean)	0.88	0.95	0.93	0.84	0.93	0.86	0.95	**0.99**	**0.99**	0.96	**0.99**
Kappa (Standard Dev.)	0.006	0.002	0.004	0.004	0.004	0.003	0.009	0.004	0.004	0.005	0.002

on the performance of the proposed SBDSM method. In this analysis experiments, the number of selected training and test samples for the Indian Pines, Salinas, University of Pavia, and Washington DC images are set to the same as that in the aforementioned comparison experiments. The experiments for the proposed algorithm are repeated over ten times with different randomly selected samples to reduce the bias induced by random sampling.

The sparsity level, i.e., K, was varied from 1 to 20. Fig. 10 shows the overall accuracies of the proposed SBDSM method under different sparsity levels for the four-test HSI images. We can observe that the overall accuracies of the SBDSM method clearly increase for all the images, as the sparsity level rises from 1 to 3. By further increasing the sparsity level from 4 to 8, the overall accuracies of the SBDSM either demonstrate a slight increase or become comparatively stable. Since larger

TABLE IV
CLASSIFICATION ACCURACIES OF THE UNIVERSITY OF PAVIA IMAGE OBTAINED BY THE SVM [2], EMP [14], LORSAL-MLL [58], PIXELWISE SRC [27], REGION-BASED SRC [27], PIXEL-LCKSVD [50], SUPERPIXEL-SVM, SBDSM-NoDL, SBDSM-KSVD [45], SBDSM-Cseg [57], AND SBDSM METHODS. CLASS-SPECIFIC ACCURACIES ARE IN PERCENTAGE

Class	SVM	EMP	LORSAL-MLL	Pixel-wise SRC	Region-based SRC	Pixel-LC KSVD	Superpixel-SVM	SBDSM-NoDL	SBDSM-KSVD	SBDSM-NCUT	SBDSM
Asphalt	97.16	**98.98**	95.38	62.20	82.72	64.75	96.26	93.85	92.86	77.57	94.79
Meadows	97.89	**98.92**	98.64	79.69	96.19	81.99	97.57	97.48	97.33	96.00	97.65
Gravel	77.34	94.45	88.04	70.30	99.22	70.61	81.48	99.35	**100**	98.65	99.70
Trees	87.88	**96.63**	98.00	91.90	92.32	92.55	76.37	87.92	90.60	88.74	90.91
Metal sheets	97.74	98.21	99.86	99.66	99.08	99.75	96.20	97.53	99.83	97.80	**99.98**
Bare soil	80.43	85.25	**99.75**	69.28	99.60	69.46	98.08	99.42	99.74	99.48	99.69
Bitumen	68.04	87.34	97.63	87.92	99.16	86.98	96.80	99.99	**100**	98.40	**100**
Bricks	85.82	98.33	92.66	71.79	96.58	73.06	98.62	98.55	**99.17**	95.66	99.05
Shadows	99.83	**99.91**	99.96	97.89	86.99	98.27	98.56	93.36	98.45	91.55	98.80
OA (Mean in %)	91.45	96.32	97.25	76.58	94.23	78.13	95.27	96.64	96.88	93.13	**97.33**
OA (Standard Dev.)	0.44	0.42	0.23	0.70	0.42	0.51	0.79	0.53	0.64	0.46	0.33
AA (Mean in %)	88.01	95.33	96.29	81.18	94.58	81.94	93.33	96.38	97.55	93.76	**97.84**
AA (Standard Dev.)	0.431	0.43	0.35	0.57	0.28	0.37	0.65	0.46	0.40	0.43	0.29
Kappa (Mean)	0.88	0.95	**0.96**	0.69	0.92	0.71	0.94	0.95	**0.96**	0.91	**0.96**
Kappa (Standard Dev.)	0.005	0.005	0.002	0.008	0.005	0.006	0.009	0.007	0.008	0.006	0.005

TABLE V
CLASSIFICATION OF THE WASHINGTON DC IMAGE OBTAINED BY THE SVM [2], EMP [14], LORSAL-MLL [58], PIXELWISE SRC [27], REGION-BASED SRC [27], PIXEL-LCKSVD [50], SUPERPIXEL-SVM, SBDSM-NoDL, SBDSM-KSVD [45], SBDSM-Cseg [57], AND SBDSM METHODS. CLASS-SPECIFIC ACCURACIES ARE IN PERCENTAGE

Class	SVM	EMP	LORSAL-MLL	Pixel-Wise-SRC	Region-based SRC	Pixel-LC KSVD	Superpixel-SVM	SBDSM-NoDL	SBDSM-KSVD	SBDSM-Cseg	SBDSM
Roof	91.97	95.39	95.13	90.95	96.20	91.44	**97.14**	95.78	96.42	90.51	96.38
Road	97.21	98.69	98.70	96.93	95.03	97.78	98.50	97.74	97.69	94.63	**98.92**
Trail	83.61	88.80	91.79	82.82	**95.70**	87.07	93.83	93.77	94.20	91.99	95.19
Grass	97.32	**99.63**	97.73	96.82	88.44	97.47	96.22	97.42	98.42	93.07	98.24
Shadow	98.20	98.53	97.78	94.55	93.12	94.84	**98.54**	95.25	95.18	92.40	96.98
Tree	98.89	**99.16**	97.50	96.57	85.90	97.93	88.20	93.82	94.13	86.37	95.09
OA (Mean in %)	93.80	96.31	96.23	92.70	93.39	93.88	95.96	95.77	96.18	91.55	**96.84**
OA (Standard Dev.)	1.08	0.41	0.34	0.62	0.53	0.47	0.78	1.04	0.75	0.46	0.57
AA (Mean in %)	94.53	96.70	96.45	93.11	92.40	94.42	95.22	95.63	96.01	91.49	**96.80**
AA (Standard Dev.)	0.92	0.41	0.43	0.61	0.56	0.53	0.49	1.10	0.73	0.40	0.57
Kappa (Mean)	0.92	0.95	0.95	0.91	0.92	0.92	0.94	0.95	0.95	0.90	**0.96**
Kappa (Standard Dev.)	0.013	0.005	0.004	0.006	0.006	0.006	0.009	0.013	0.009	0.005	0.007

TABLE VI
AVERAGE RUN TIME (SECONDS) OVER TEN REALIZATIONS FOR THE CLASSIFICATION OF THE FOUR-TEST HSI IMAGES BY THE SVM [2], EMP [14], LORSAL-MLL [58], PIXELWISE SRC [27], REGION-BASED SRC [27], PIXEL-LCKSVD [50], SUPERPIXEL-SVM, SBDSM-NoDL, SBDSM-KSVD [45], SBDSM-CSEG [57], AND SBDSM METHODS

IMAGES	SVM	EMP	LORSAL-MLL	Pixel-Wise SRC	Region-based SRC	Pixel-LCKSVD	Superpixel-SVM	SBDSM-NoDL	SBDSM-KSVD	SBDSM-Cseg	SBDSM
Indian Pines	156.2	61.5	68.6	16.4	92.6	22.1	156.8	8.7	7.9	18.0	**6.1**
Salinas	45.8	18.6	89.9	80.8	398.2	18.6	47.1	15.8	7.6	42.7	**6.7**
U. Pavia	176.6	106.3	383.8	57.5	490.0	180.4	178.3	38.2	28.9	130.1	**25.3**
Washington DC	20.3	8.9	36.9	9.5	25.8	11.9	20.8	13.1	8.8	72.0	**7.8**

sparsity levels will create more computational cost in the SOMP algorithm [41], the sparsity level K is selected as 3 in the aforementioned experiments. In addition, when the sparsity level further increases from 9 to 20, the performance of the SBDSM method generally degrades for the Indian Pines, Salinas, and Washington DC images, whereas its performance for the University of Pavia image remains stable. That is because the selection of the sparsity level should be related to the number of dictionary atoms in each class. For the Indian Pines image, since the number of dictionary atoms for some classes is very small, larger sparsity levels will enable the SBDSM algorithm to select more atoms from other incorrect classes, consequently deteriorating the classification accuracy. In contrast, for the University of Pavia image, the number of dictionary atoms within each class is large (e.g., 250), and thus, the performance of the SBDSM exhibits little variations for sparsity levels ranging from 1 to 20.

The subdictionary size per class N_r, varied from 10% to 100% of the number of training samples in each class. Fig. 11 shows the overall accuracies of the proposed SBDSM method under different subdictionary sizes on the four HSI images. As can be observed, the performances of the proposed SBDSM

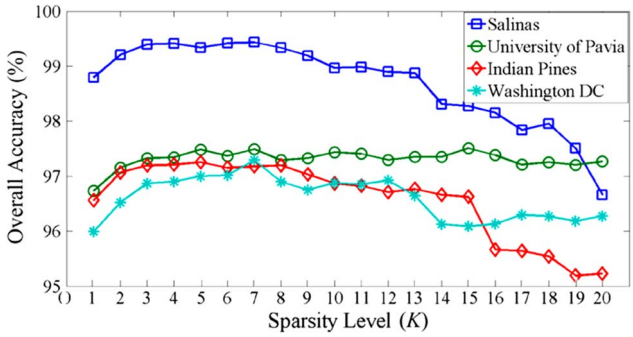


Fig. 10. Effect of the sparsity level on the proposed SBDSM algorithm for the four HSI images.

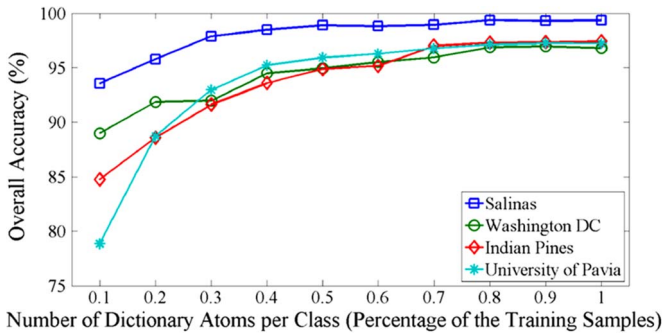


Fig. 11. Effect of the dictionary size on the proposed SBDSM algorithm for the four HSI images.

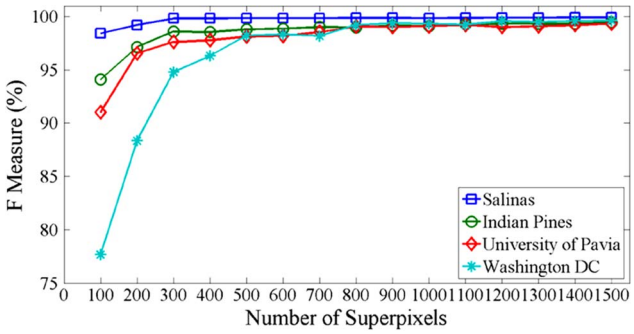


Fig. 12. Effect of the number of superpixels L on the segmentation results (F measure) on the four HSI images.

method are comparatively stable when the dictionary size goes from 80% to 100% of the number of the training samples. Therefore, 80% of the number of the training samples is chosen as the dictionary size used in our experiments. Furthermore, as the dictionary size decreases from 70% to 40% of the number of the training samples, the overall accuracies of the proposed SBDSM method slightly decreases. It should be also noted that when the dictionary size reaches about 20% of the number of the training samples, the overall accuracies of the proposed SBDSM method for all the test images are higher than 87%.

The number of superpixels L was varied within the range $[100, \ldots, 1500]$. First, the F-measure was adopted to examine the performance of the oversegmentation method [40] in order to create the superpixel map for different numbers of superpixels, as shown in Fig. 12. As the number of superpixels initially

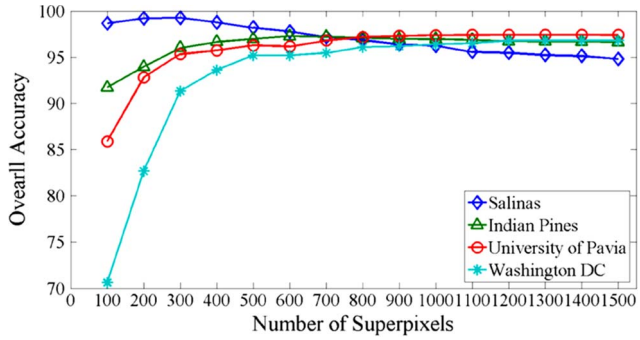


Fig. 13. Effect of the number of superpixels L on the proposed SBDSM algorithm on the four HSI images.

increases, i.e., for relatively few superpixels, the segmentation accuracies (F-measures) for the four images generally increase. Then, from the figure, it can be seen that optimal segmentation accuracies will be approximately obtained, when the number of superpixel reaches 300, 600, 1000, 1300 for the Salinas, Indian Pines, University of Pavia, Washington DC, respectively. As can be seen, although the Salinas image is large, it requires a smaller number (300) of superpixels, as compared with numbers needed for the other three images. That is because large homogenous regions do exist in the Salinas image. Therefore, that image can be easily segmented. In addition, the corresponding classification results (OA) of the proposed SBDSM for a different number of superpixels are also illustrated in the Fig. 13. We can see that the performance of the proposed SBDSM method improves as the superpixel number L increases from 100 to 600 for the Indian Pines image, 100 to 300 for the Salinas image, 100 to 1000 for the University of Pavia image and 100 to 1300 for the Washington DC image. This trend in the classification accuracies generally corresponds to that on the segmentation results. However, as the number of superpixels further increases, the performance of the proposed SBDSM method degrades for the Salinas image. The reason is that if the number of superpixels L is too large, the size of each superpixel will become very small, and thus, the spatial information for large homogenous regions in the Salinas image will not be sufficiently exploited in classification.

F. Classification Results With Different Number of Training Samples

Here, it is examined how the number of training samples affects the performance for various sparse representation-based algorithms on the Indian Pines, Salinas, University of Pavia, and Washington DC images. The parameters for all the algorithms are fixed to be the same as that in the aforementioned comparison experiments. For each class in the Indian Pines, Salinas, and Washington DC images, labeled samples were randomly selected in varied percentages (from 2.5% to 50% for Indian Pines, from 0.25% to 5% for Salinas, and from 1% to 10%) as the training samples, and the remaining samples for the test. For the University of Pavia image, a balanced training set was constructed by randomly choosing from 100 to 500 pixels for each class.

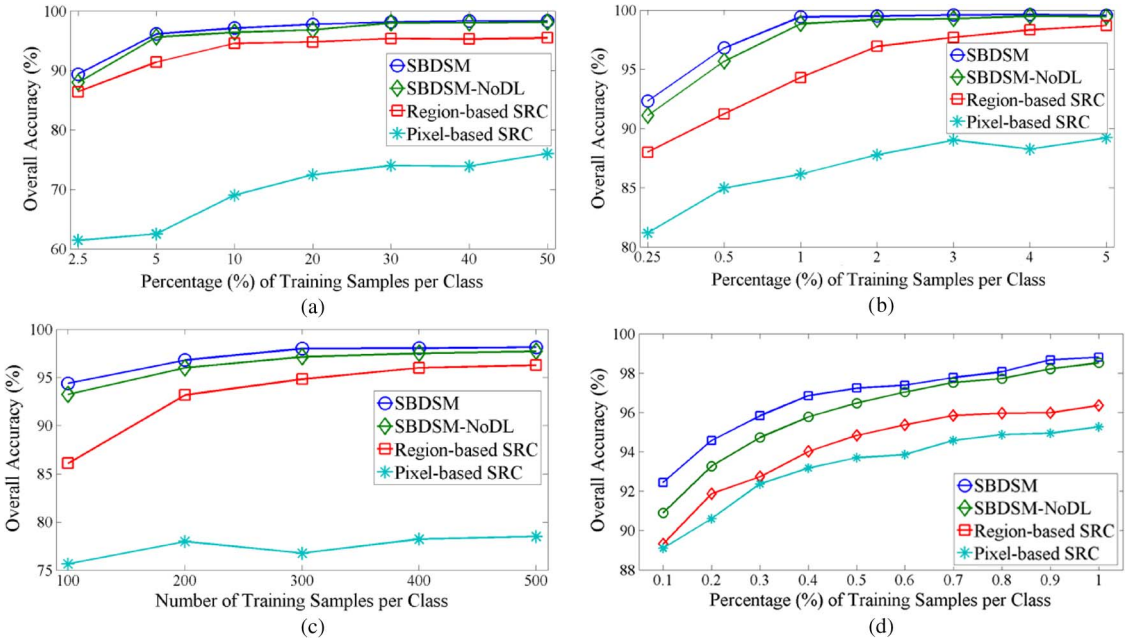


Fig. 14. Effect of the number of training samples on Pixelwise SRC, Region-based SRC, SBDSM-NoDL, and SBDSM for the (a) Indian Pines Image, (b) Salinas Image, (c) University of Pavia Image, and (d) Washington DC Image.

The classification accuracies (OA), which are averaged over five runs for each classification approach at each different number of training samples are shown in Fig. 14. As can be observed, in most of the cases, the accuracies for all the classifier monotonically increase as the number of training samples increase. In addition, it can be seen that the SBDSM classifier generally outperforms all other sparse representation-based classifiers in terms of accuracies, particularly when a limited number of training samples is available.

V. Conclusion

In this paper, a novel SBDSM has been proposed for HSI classification. The proposed SBDSM can classify one superpixel at a time and, thus, is more efficient than other sparsity-based approaches for HSI classification. In addition, the size and shape of each superpixel can be adaptively changed according to the spatial structures of the HIS, and therefore, the spatial contexts can be effectively exploited. Furthermore, by utilizing the class label information for both training samples and dictionary atoms, a class-labeled OMP algorithm for the discriminative K-SVD learning algorithm is proposed, which can efficiently train both a discriminative dictionary and a classifier. The results of the experiments in this paper demonstrate that the proposed SBDSM algorithm outperforms several well-known classification approaches in terms of both classification accuracies and computational speed.

In the experiments, the number of superpixels was selected empirically for different images. Therefore, one of our ongoing research directions is to establish a more systematic way of adaptively selecting the number of superpixels for different conditions. In addition, there is a strong incentive to apply the superpixel-based sparse representation model to other remote sensing applications (e.g., denoising, change detection, and object recognition).

Acknowledgment

The authors would like to thank Dr. J. Li and Dr. M.-Y. Liu for providing the software for the LORSAL-MLL and oversegmentation method, respectively, on their websites.[2,3]

References

[1] V. Vapnik, *The Nature of Statistical Learning Theory*. Berlin, Germany: Springer-Verlag, 1995.
[2] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.
[3] M. Chi and L. Bruzzone, "Semisupervised classification of hyperspectral images by SVMs optimized in the primal," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1870–1880, Jun. 2007.
[4] D. Böhning, "Multinomial logistic regression algorithm," *Ann. Inst. Statist. Math.*, vol. 44, no. 1, pp. 197–200, Mar. 1992.
[5] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 809–823, Mar. 2012.
[6] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4085–4098, Nov. 2010.
[7] Y. Zhong and L. Zhang, "An adaptive artificial immune network for supervised classification of multi-/hyperspectral remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 894–909, Mar. 2012.
[8] H. Jiao, Y. Zhong, and L. Zhang, "Artificial DNA computing-based spectral encoding and matching algorithm for hyperspectral remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 4085–4104, Oct. 2012.
[9] J. Li, P. R. Marpu, A. Plaza, J. M. Bioucas-Dias, and J. A. Benediktsson, "Generalized composite kernel framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4816–4829, Sep. 2013.

[2]http://www.lx.it.pt/~jun/
[3]https://sites.google.com/site/seanmingyuliu/

[10] Y. Zhong, Y. Lin, and L. Zhang, "A support vector conditional random classifier with a Mahalanobis distance boundary constraint for high spatial resolution remotes sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1314–1330, Apr. 2014.

[11] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral-spatial classification of hyperspectral imagery based on partitional clustering techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2973–2987, Aug. 2009.

[12] S. Prasad, M. Cui, W. Li, and J. E. Fowler, "Segmented mixture-of-Gaussian classification for hyperspectral image analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 138–142, Jan. 2014.

[13] L. Zhang, Y. Zhong, B. Huang, J. Gong, and P. Li, "Dimensionality reduction based on clonal selection for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 4172–4186, Dec. 2007.

[14] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.

[15] L. Zhang, L. Zhang, D. Tao, and X. Huang, "Tensor discriminative locality alignment for hyperspectral image spectral-spatial feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 242–256, Jan. 2013.

[16] L. Zhang, L. Zhang, D. Tao, and X. Huang, "On combining multiple features for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 879–893, Mar. 2012.

[17] J. Li, H. Zhang, L. Zhang, X. Huang, and L. Zhang, "Joint collaborative representation with multi-task learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5923–5936, Sep. 2014.

[18] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, Feb. 2009.

[19] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[20] J. Wright *et al.*, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.

[21] S. Li, L. Fang, and H. Yin, "An efficient dictionary learning algorithm and its application to 3-D medical image denoising," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 2, pp. 417–427 Feb. 2012.

[22] S. Li, H. Yin, and L. Fang, "Remote sensing image fusion via sparse representations over learned dictionaries," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4779–4789, Sep. 2013.

[23] Y. Qian and M. Ye, "Hyperspectral imagery restoration using nonlocal spectral-spatial structured sparse representation with noise estimation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 499–515, Apr. 2013.

[24] Z. Xing, M. Zhou, A. Castrodad, G. Sapiro, and L. Carin, "Dictionary learning for noisy and incomplete hyperspectral images," *SIAM J. Imag. Sci.*, vol. 5, no. 1, pp. 33–56, Jan. 2012.

[25] L. Fang *et al.*, "Fast acquisition and reconstruction of optical coherence tomography images via sparse representation," *IEEE Trans. Med. Imag.*, vol. 32, no. 11, pp. 2034–2049, Nov. 2013.

[26] L. Fang, S. Li, Q. Nie, J. A. Izatt, C. A. Toth, and S. Farsiu, "Sparsity based denoising of spectral domain optical coherence tomography images," *Biomed. Opt. Exp.*, vol. 3, no. 5, pp. 927–942, May 2012.

[27] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.

[28] U. Srinivas, Y. Chen, V. Monga, N. M. Nasrabadi, and T. D. Tran, "Exploiting sparsity in hyperspectral image classification via graphical models," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 3, pp. 505–509, May 2013.

[29] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification via kernel sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 217–231, Jan. 2013.

[30] H. Zhang, J. Li, Y. Huang, and L. Zhang, "A nonlocal weighted joint sparse representation classification method for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2057–2066, Jun. 2014.

[31] J. Li, H. Zhang, Y. Huang, and L. Zhang, "Hyperspectral image classification by nonlocal joint collaborative representation with a locally adaptive dictionary," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3707–3719, Jun. 2014.

[32] W. Li, E. W. Tramel, S. Prasad, and J. E. Fowler, "Nearest regularized subspace for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 477–489, Jan. 2012.

[33] Y. Y. Tang, H. Yuan, and L. Li, "Manifold-based sparse representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7606–7618, Dec. 2014.

[34] H. Yuan, Y. Tang, Y. Lu, L. Yang, and H. Luo, "Hyperspectral image classification based on regularized sparse representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2174–2182, Jun. 2014.

[35] Z. Wang, N. M. Nasrabadi, and T. S. Huang, "Spatial-spectral classification of hyperspectral images using discriminative dictionary designed by learning vector quantization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4808–4822, Aug. 2013.

[36] J. Li, H. Zhang and L. Zhang, "Column-generation kernel nonlocal joint collaborative representation for hyperspectral image classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 94, no. 8, pp. 25–36, Aug. 2014.

[37] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral-spatial hyperspectral Image classification via multiscale adaptive sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7738–7749, Dec. 2014.

[38] G. Mori, X. Ren, A. A. Efros, and J. Malik, "Recovering human body configurations: Combining segmentation and recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2004, pp. II-326–II-333.

[39] R. Achanta *et al.*, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2281, Nov. 2012.

[40] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 2097–2104.

[41] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, Mar. 2006.

[42] G. Zhang, X. Jia, and N. M. Kwok, "Super pixel based remote sensing image classification with histogram descriptors on spectral and spatial data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2012, pp. 4335–4338.

[43] J. Liu, W. Yang, S. Tan, and Z. Wang, "Remote sensing image classification based on random projection super-pixel segmentation," in *Proc. SPIE*, 2013, 89210T.

[44] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proc. IEEE*, vol. 98, no. 6, pp. 1045–1057, Jun. 2010.

[45] Q. Zhang and B. Li, "Discriminative K-SVD for dictionary learning in face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 2691–2698.

[46] M. Aharon, M. Elad, and A. M. Bruckstein, "The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[47] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[48] R. Rubinstein, M. Zibulevsky, and M. Elad, "Double sparsity: Learning sparse dictionaries for sparse signal approximation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1553–1564, Mar. 2010.

[49] Z. Jiang, Z. Lin, and L. S. Davis, "Learning a discriminative dictionary for sparse coding via label consistent K-SVD," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 1697–1704.

[50] Z. Jiang, Z. Lin, and L. Davis, "Label consistent K-SVD: Learning a discriminative dictionary for recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2651–2664, Nov. 2013.

[51] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *J. Constructive Approx.*, vol. 13, no. 1, pp. 57–98, Mar. 1997.

[52] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Trans. Image Process.*, vol. 17, no. 1, pp. 53–69, Jan. 2008.

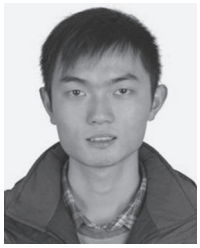
[53] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[54] I. Jolliffe, *Principal Component Analysis*. Hoboken, NJ, USA: Wiley, 2005.

[55] L. Zhang, L. Zhang, D. Tao, and X. Huang, "Sparse transfer manifold embedding for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1030–1043, Feb. 2014.

[56] Y. Qian, M. Ye, and J. Zhou, "Hyperspectral image classification based on structured sparse logistic regression and 3-D wavelet texture feature," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2276–2291, Apr. 2013.

[57] X. Ren and J. Malik, "Learning a classification model for segmentation," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, 2003, pp. 10–17.
[58] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Hyperspectral image segmentation using a new Bayesian approach with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3947–3960, Oct. 2011.
[59] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, Apr. 2011.
[60] J. A. Gualtieri and R. F. Cromp, "Support vector machines for hyperspectral remote sensing classification," in *Proc. SPIE*, 1999, pp. 221–232.
[61] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. S110–S122, Sep. 2009.
[62] A. Levinshtein, C. Sminchisescu, and S. Dickinson, "Optimal image and video closure by superpixel grouping," *Int. J. Comput. Vis.*, vol. 100, no. 1, pp. 99–119, Oct. 2012.
[63] B. Song *et al.*, "Remotely sensed image classification using sparse representations of morphological attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 5122–5136, Aug. 2014.
[64] R. Ji *et al.*, "Spectral-spatial constraint hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 3, pp. 1811–1824, Mar. 2014.



Leyuan Fang (S'10–M'14) received the B.S. degree in electrical engineering from the Hunan University of Science and Technology, Xiangtan, China, in 2008, and the Ph.D. degree from the College of Electrical and Information Engineering, Hunan University, Changsha, China, in 2008.

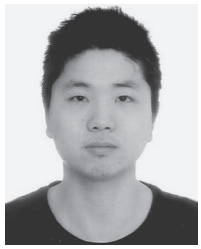
Since September 2011, he has been a Visiting Ph.D. Student with the Department of Ophthalmology, Duke University, Durham, NC, USA, supported by the China Scholarship Council. His research interests include sparse representation and multiresolution analysis in remote sensing and medical image processing.

Mr. Fang was the recipient of the Scholarship Award for Excellent Doctoral Student granted by Chinese Ministry of Education in 2011.

Shutao Li (M'07) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Hunan University, Changsha, China, in 1995, 1997, and 2001, respectively.

From May 2001 to October 2001, he joined the College of Electrical and Information Engineering, Hunan University, in 2001. He was a Research Associate with the Department of Computer Science, Hong Kong University of Science and Technology, Kowloon, Hong Kong. From November 2002 to November 2003, he was a Postdoctoral Fellow with the Royal Holloway College, University of London, London, U.K., working with Prof. J. Shawe-Taylor. During April 2005 to June 2005, he visited the Department of Computer Science, Hong Kong University of Science and Technology as a Visiting Professor. Now, he is a Full Professor with the College of Electrical and Information Engineering, Hunan University. He has authored or coauthored more than 160 refereed papers. His professional interests are compressive sensing, sparse representation, image processing, and pattern recognition.

Dr. Li is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, a member of the Editorial Board of the journal *Information Fusion* and the *Sensing and Imaging*. He was the recipient of two Second-Grade National Awards at the Science and Technology Progress of China in 2004 and 2006.



Xudong Kang (S'13) received the B.Sc degree from Northeast University, Shenyang, China, in 2007. He is currently working toward the Ph.D. degree in electrical engineering at Hunan University, Changsha, China.

In 2012–2013, he was with the University of Iceland, Reykjavik, Iceland, as a Visiting Ph.D. student in electrical engineering. He is engaged in image fusion, image super-resolution, pansharpening, and hyperspectral image classification.

Jón Atli Benediktsson (S'84–M'90–SM'99–F'04) received the Cand.Sci. degree in electrical engineering from the University of Iceland, Reykjavik, in 1984 and the M.S.E.E. and Ph.D. degrees from Purdue University, West Lafayette, IN, USA, in 1987 and 1990, respectively.

He is currently Pro Rector for Academic Affairs and Professor of electrical and computer engineering with the University of Iceland. His research interests are in remote sensing, biomedical analysis of signals, pattern recognition, image processing, and signal processing, and he has published extensively in those fields.

Prof. Benediktsson was the 2011–2012 President of the IEEE Geoscience and Remote Sensing Society (GRSS) and has been on the GRSS Administrative Committee since 2000. He was Editor-in-Chief of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS) from 2003 to 2008 and has served as Associate Editor of TGRS since 1999, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS since 2003 and IEEE ACCESS since 2013. He is on the Editorial Board of the PROCEEDINGS OF THE IEEE, the International Editorial Board of the *International Journal of Image and Data Fusion* and was the Chairman of the Steering Committee of *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (J-STARS) from 2007–2010. He is a Cofounder of the biomedical start-up company Oxymap. He is a Fellow of the Society of Photographic Instrumentation Engineers. He is a member of the 2014 IEEE Fellow Committee. He was the recipient of the Stevan J. Kristof Award from Purdue University in 1991 as an outstanding graduate student in remote sensing. In 1997, he was the recipient of the Icelandic Research Council's Outstanding Young Researcher Award, in 2000, he was granted the IEEE Third Millennium Medal, in 2004, he was a corecipient of the University of Iceland's Technology Innovation Award, in 2006 he was the recipient of the yearly research award from the Engineering Research Institute of the University of Iceland, and in 2007, he received the Outstanding Service Award from the IEEE Geoscience and Remote Sensing Society. He was corecipient of the 2012 IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING Paper Award and in 2013 he was corecipient of the IEEE GRSS Highest Impact Paper Award. In 2013, he was the recipient of the IEEE/VFI Electrical Engineer of the Year Award. He is a member of the Association of Chartered Engineers in Iceland (VFI), Societas Scinetiarum Islandica, and Tau Beta Pi.