

Research Article

Dictionary-Based, Clustered Sparse Representation for Hyperspectral Image Classification

Zhen-*tao* Qin,^{1,2} Wu-*nian* Yang,¹ Ru Yang,² Xiang-*yu* Zhao,² and Teng-*jiao* Yang²

¹Key Laboratory of Geo-Special Information Technology, Ministry of Land and Resources, Institute of Remote Sensing & GIS, Chengdu University of Technology, Chengdu, Sichuan 610059, China

²Panzhuhua College, Panzhuhua, Sichuan 617000, China

Correspondence should be addressed to Zhen-*tao* Qin; 309507443@qq.com and Wu-*nian* Yang; ywn@cdut.edu.cn

Received 3 November 2014; Revised 14 December 2014; Accepted 26 December 2014

Academic Editor: Shinichi Morita

Copyright © 2015 Zhen-*tao* Qin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents a new, dictionary-based method for hyperspectral image classification, which incorporates both spectral and contextual characteristics of a sample clustered to obtain a dictionary of each pixel. The resulting pixels display a common sparsity pattern in identical clustered groups. We calculated the image's sparse coefficients using the dictionary approach, which generated the sparse representation features of the remote sensing images. The sparse coefficients are then used to classify the hyperspectral images via a linear SVM. Experiments show that our proposed method of dictionary-based, clustered sparse coefficients can create better representations of hyperspectral images, with a greater overall accuracy and a Kappa coefficient.

1. Introduction

With the development of electronic spectrum theory as well as electronic and computer technology, researchers have developed hyperspectral remote sensing (HRS) at full speed in recent years. The resulting hyperspectral remote sensing images are of a much better quality, which entails the need for larger storage capacity. Hyperspectral imagery (HI) captures detailed terrestrial information with high resolution in both the spatial and spectral dimensions [1]. Analyzing hyperspectral remote sensing data can yield abundant spectral information and detailed features [2]. Substances that cannot normally be detected by multispectral remote sensing technique can be successfully studied using hyperspectral remote sensing. HRS data has become an important data source in fields such as precision agriculture, natural disaster research, atmospheric observation, fog monitoring, environmental monitoring, and resource investigation [3].

Hyperspectral remote sensing images are composed of pixels that reflect the characteristics of the terrain object. Each pixel represents the surface features of hundreds of wavelengths of solar radiation. In mathematics, these pixels are modeled as members of a vector space. However, due

to the inherent limitations of hyperspectral sensors, which often neglect the correlation between the signals, sample data indicates rates that far exceed the effective dimension of the signals, thereby causing issues with dimensionality. In recent years, researchers have proposed many dimension reduction methods, such as principal component analysis (PCA) [4], linear discriminant analysis (LDA) [5], and independent component analysis (ICA) [6]. Since 2005, Camps-Valls and Bruzzone have carried on the research of hyperspectral remote sensing image classification using the machine learning method [7]. Understanding the causes of signal representation in a low-dimensional model is a recent research trend, known as dictionary learning [8]. The idea of dictionary learning is to represent a signal using a linear combination of a few elements from a dictionary, which is taken from the data. Each data point is represented through a sparse vector of coefficients, as a member of a low-dimensional subspace, spanned by a few dictionary elements. Iordache et al. studied the HI unmixing using the dictionary learning method [9]. In 2010, Charles et al. achieved the sparse representation of hyperspectral remote sensing images using the common dictionary learning method for images [10]. Then, in 2013, Soltani-Farani et al. created a remote sensing image sparse

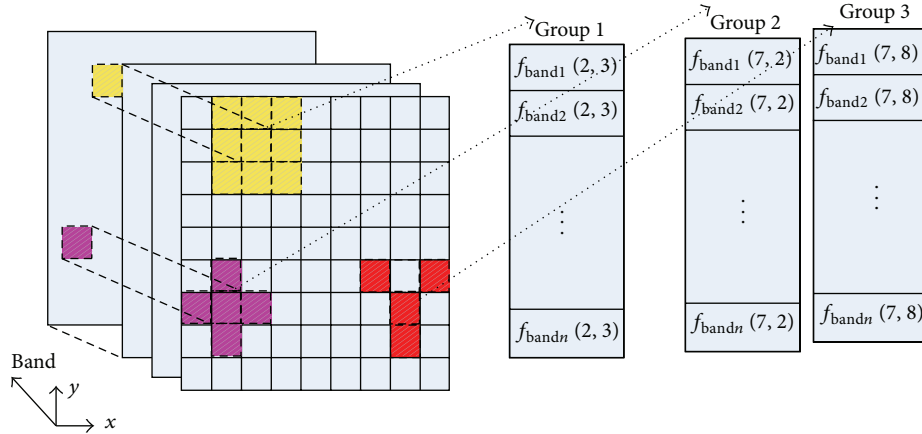


FIGURE 1: The pixels of a hyperspectral image partitioned into a number of different clustered groups.

representation in a spatial domain using the spatial relationships within the hyperspectral remote sensing image [11].

In this paper we propose a new high-quality and efficient classification technique that extends existing dictionary learning-based classification frameworks in several aspects. We incorporate both the spectral and contextual characteristics of a hyperspectral sample by clustering remote sensing images to obtain a dictionary of pixels and present a cluster-based dictionary learning method; pixels that belong to the same cluster group are often made up of the same materials. This property holds for various hyperspectral image singularities such as straight and corner edges, as shown in Figure 1. Using a linear SVM as a classifier, we completely classified the hyperspectral remote sensing images. We compare this cluster-based dictionary learning method with other alternatives for classification and show that it performs significantly better in terms of both accuracy and a Kappa coefficient.

2. Basic Model

For a set of pixels of hyperspectral image, let $\mathbf{x} \in \mathbf{R}^M$, and the fundamental goal of the dictionary learning method is to find a set of atomic signals $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_K]$ to represent the hyperspectral data by a small number of terms in a linear generative model; that is,

$$\mathbf{X} = \mathbf{D}\mathbf{y} + \boldsymbol{\varepsilon}. \quad (1)$$

In this paper, we use lowercase letters to represent vectors (such as \mathbf{x}) and capital letters to represent matrices. Moreover, $\boldsymbol{\varepsilon}$ is a small residual due to modeling \mathbf{x} in a linear manner with the sparse representation vector $\mathbf{y} \in \mathbf{R}^K$. The formulation of (1) is often a regularized least squares optimization as follows:

$$\arg \min \frac{1}{2} \|\mathbf{X} - \mathbf{D}\mathbf{Y}\|_F^2 + \gamma S(\mathbf{Y}), \quad (2)$$

where $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]$, $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_N]$, and $\|\cdot\|_F$ denotes the Frobenius norm of a matrix. The second part of γ is a parameter that trades off between the data fidelity (least-squares) term and the sparsity based regularizer (the l_1

norm); $S(\mathbf{Y})$ can be interpreted as finding a maximum a posteriori (MAP) estimate of the coefficients under the assumptions of Gaussian noise and a prior independent identically distributed (i.i.d.); traditionally, a Laplacian distribution is preferred as it leads to the well-known Lasso or ℓ_1 minimization and can be expressed as [12]

$$\arg \min_{\mathbf{y}_i} \frac{1}{2} \|\mathbf{X} - \mathbf{D}\mathbf{Y}\|_F^2 + \gamma \sum_{i=1}^N \|\mathbf{y}_i\|_1. \quad (3)$$

Tibshirani et al. used (3) to solve problem (2). The basic problem of dictionary learning is to learn through sparse regularization representations $\mathbf{y}_1, \dots, \mathbf{y}_N$. Its vectors form $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(K)}$. The above optimization is convex in either \mathbf{D} or \mathbf{Y} , but not in both. Commonly, a two-step strategy is used for this problem.

(1) *Sparse Coding*. In this step, \mathbf{D} is fixed and the optimization is solved with regard to \mathbf{Y} . The objective function (3) is converted to (4) and can be solved for each \mathbf{y}_i independently; that is,

$$\arg \min_{\mathbf{y}_i} \frac{1}{2} \|\mathbf{x}_i - \mathbf{D}\mathbf{y}_i\|_F^2 + \gamma \|\mathbf{y}_i\|_1. \quad (4)$$

Several efficient algorithms have been proposed to solve this issue (4) in recent years, though this paper implements the data [11] provided by the SPAMS toolbox.

(2) *Dictionary Update*. In this step, researchers apply the dictionary update, in which \mathbf{Y} is fixed and the optimization becomes

$$\begin{aligned} \arg \min_{\mathbf{D}} \quad & \frac{1}{2} \|\mathbf{X} - \mathbf{D}\mathbf{Y}\|_F^2 \\ \text{s.t.} \quad & \forall i \quad \|\mathbf{d}_i\|_2 \leq 1, \end{aligned} \quad (5)$$

which is quadratic in \mathbf{D} . The gradient of the objective function equals $\mathbf{D}\mathbf{Y}\mathbf{Y}^T - \mathbf{X}\mathbf{Y}^T$, in which zero is used for $\mathbf{D} = \mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T)^{-1}$. There are many ways of solving the problem; we use the block coordinate descent (BCD) [13], which updates

the dictionary atoms iteratively. Since the objective function is strongly convex, BCD is guaranteed to achieve the unique solution. The atom of j of objective function can expressed as $(1/2)\|\mathbf{R}_j - \mathbf{d}\mathbf{Y}_T^{jT}\|_F^2$, where \mathbf{Y}_T^{jT} is the row j of \mathbf{Y} and $\mathbf{R}_j = \mathbf{X} - \sum_{i \neq j} \mathbf{d}_i \Psi_T^{jT}$. In order to solve for \mathbf{d}_j , the algebra is as follows:

$$\begin{aligned} \arg \min_{\mathbf{D}} \quad & \frac{1}{2} \|\mathbf{X} - \mathbf{D}\mathbf{Y}\|_F^2 \\ \text{s.t.} \quad & \forall i \quad \|\mathbf{d}_i\|_2 \leq 1. \end{aligned} \quad (6)$$

3. Clustered Sparse Representation for Hyperspectral Image Classification

Recently, Song and Jiao used the sparse representation method for hyperspectral image classification [14], in which the sparse representation coefficients \mathbf{y}_i are considered to be independent of each other. Soltani-Farani et al. [11] partitioned the pixels into groups of the same size, such as group 1 in Figure 1. Yet, the features of the group are not necessarily similar when grouped by identical size, such as those in groups 2 and 3 in Figure 1. In order to solve this problem and further improve the classification accuracy, we propose partitioning the pixels of the hyperspectral images into a number of spatial neighborhoods called groups by clustering. Pixels that belong to the same cluster group are often made up of the same materials, so we assume that their representations use a common set of atoms from the dictionary. Thus, the sparse representations of the HSI pixels that belong to the same group are no longer independent. In fact, the pixels in the same cluster groups have revealed hidden relationships within the spectral bands. HSI are a collection of hundreds of images that have been acquired simultaneously in narrow and adjacent spectral bands. In this research, $\mathbf{x}_1, \dots, \mathbf{x}_N$ denote the representation of the pixels in a hyperspectral image and define the cluster groups G_1, \dots, G_n as nonoverlapping image patches. Figure 1 shows how the pixels of a hyperspectral image may be partitioned into a number of different groups. Accounting for the above assumption, the establishment of a sparse representation model can now be written as

$$\mathbf{X}_{G_i} = \mathbf{D}\mathbf{Y}_{G_i} + \mathbf{E}_{G_i}. \quad (7)$$

In this model, the columns of \mathbf{Y}_{G_i} and \mathbf{E}_{G_i} are the sparse representations and error vectors corresponding to the hyperspectral samples, respectively. In order to get the dictionary and sparse representations, we employ the ℓ_2/ℓ_1 convex joint sparsity-inducing regularizer in order (2) to arrive at

$$\begin{aligned} \arg \min_{\mathbf{D}, \mathbf{Y}} \quad & \frac{1}{2} \|\mathbf{X} - \mathbf{D}\mathbf{Y}\|_F^2 + \sum_{i=1}^g \gamma_{G_i} \|\mathbf{Y}_{G_i}\|_{2,1} \\ \text{s.t.} \quad & \forall i \quad \|\mathbf{d}_i\|_2 \leq 1, \end{aligned} \quad (8)$$

where γ_{G_i} is the regularization parameter for the group of i and $\|\mathbf{Y}_{G_i}\|_{2,1}$ is the ℓ_2/ℓ_1 norm of the row of \mathbf{Y} . In order to solve the problem, we have empirically adopted a regularized M-FOCUSS algorithm [15]. By estimating each row of the ℓ_2

norm, we can update γ_G according to the estimated value. Setting the gradient of the objective function at (8) zero, we arrive at

$$\mathbf{A}\mathbf{D}^T\mathbf{D}\mathbf{Y}_G - \mathbf{A}\mathbf{X}_G^T\mathbf{D} + \gamma_G\mathbf{Y}_G = 0, \quad (9)$$

where $\mathbf{A} = \text{diag}(\|\mathbf{Y}_{G,T}^i\|_2)$. By solving (9), we arrive at $\mathbf{Y}_G = \mathbf{A}\mathbf{D}^T(\mathbf{D}\mathbf{A}\mathbf{D}^T + \gamma_G\mathbf{I})\mathbf{X}_G$. According to the relationship of the pixels in the same cluster groups have a hidden relationships within the spectral bands, showed in Figure 1. In order to implement the dictionary method, we used $\mathbf{x}_1, \dots, \mathbf{x}_N$ to denote the spectral representation of the training data with respective labels $\mathbf{I}_1, \dots, \mathbf{I}_N$ and then applied the dictionary learning formulation of (4) to these samples to yield the corresponding sparse representations $\mathbf{y}_1, \dots, \mathbf{y}_N$ and the dictionary, \mathbf{D} . When there is a new hyperspectral sampling, sparse coding can be applied (as in (4)) to find the corresponding sparse representation \mathbf{y} , which is then classified using the trained linear SVM to find the corresponding label \mathbf{I} . The specific steps are as follows:

- (1) cluster the hyperspectral image into different groups by k -means++ [16];
- (2) apply the dictionary learning method using the SPAMS toolbox to solve the formula of (4), which yields the dictionary, \mathbf{D} , and the corresponding sparse representations coefficient, \mathbf{y} , with respective labels $\mathbf{I}_1, \dots, \mathbf{I}_N$;
- (3) a linear SVM classifier is trained on the sparse representations and their corresponding labels $\mathbf{I}'_1, \dots, \mathbf{I}'_N$;
- (4) according to the sparse representation of remote sensing images, we used a linear SVM classifier to achieve classification.

4. Experimental Results and Analysis

In this section, in order to validate and test the effectiveness of the proposed clustered dictionary-based algorithm, we provide the experimental results from two sets of real hyperspectral images. We then compare the classification accuracies of the basic SVM classification (SVM) [17], which is the hot issue accompanying artificial neural network in machine learning, and it involves any practical problems such as classification and regression estimation. In this paper, Libsvm 3.17 is used to do the experiments. Classification accuracy depends on the choice of parameters. All parameters (polynomial kernel degree d , RBF kernel parameter, regularization parameter C , the composite kernel weight, and the window width w) are obtained by fivefold cross validation. The spectral-contextual dictionary learning (SCDL) is presented by Soltani-Farani et al. [11]. In the paper, the authors partitioned the pixels into groups of the same size. The clustered dictionary learning (Cluster-DL) is presented by our team; the pixels of a hyperspectral image are partitioned into a number of different groups by k -means++. We also compare the spectral characteristics that have been gathered from the dictionary learning method, which are made up of dictionary atoms with the original spectral remote sensing images' characteristics.

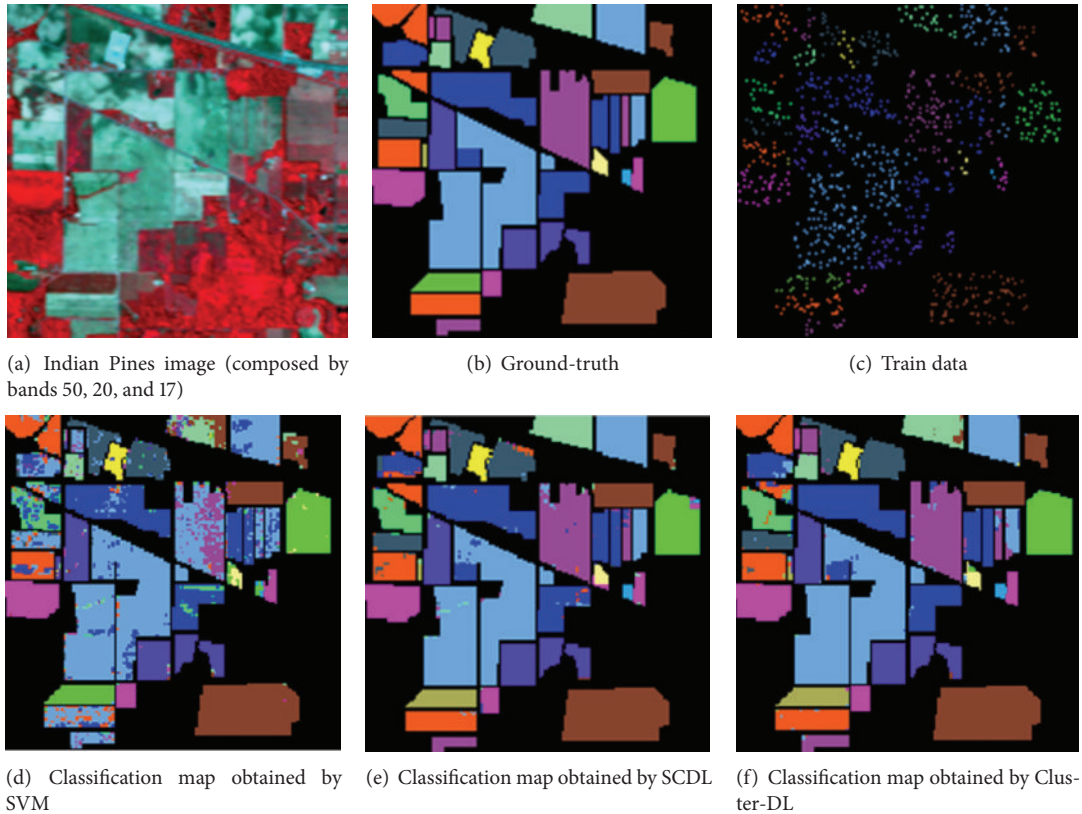


FIGURE 2: Indian Pines hyperspectral image and the comparison maps of different classification.

The experiment adopted four indicators to evaluate overall accuracy (OA), average accuracy (AA), Kappa, and execution time.

4.1. The 1st Experiment. We collected the 1st experiment over an agricultural/forested area in NW Indiana using the AVIRIS sensor, called the Indian Pines image. The image is $145 \text{ pixels} \times 145 \text{ pixels}$ and consists of 220 bands across the spectral range 0.2 to $2.4 \mu\text{m}$ and 20 noisy bands (104–108, 150–163, and 220) that correspond to the region of water absorption that has been removed. The image consists of 16 ground-truth classes; the specific classes and the number of train and test data in each class are shown in Table 1. We randomly chose 10% as the training data, as shown in Figure 2(c), and the remaining 90% is the test data. Table 2 displays the test results, which contain the OA, AA, and Kappa coefficient. The SVM classification result is shown in Figure 2(d), whereas the classification maps obtained by other methods can be found in Figures 2(e) and 2(f). As a means of visual comparison, we used learning dictionaries with 138 atoms (using 10% of the Indian Pines training data). Figure 3 demonstrates the comparison map of sample spectra for Alfalfa in the Indian Pines dataset and the learning dictionary atom obtained by SCDL and Cluster-DL.

In Figure 3, we can see the sample spectra for Alfalfa in the Indian Pines dataset, and the learning dictionary atoms obtained by Cluster-DL and SCDL are close to each sample. Relatively speaking, the Cluster-DL is closer to the sample

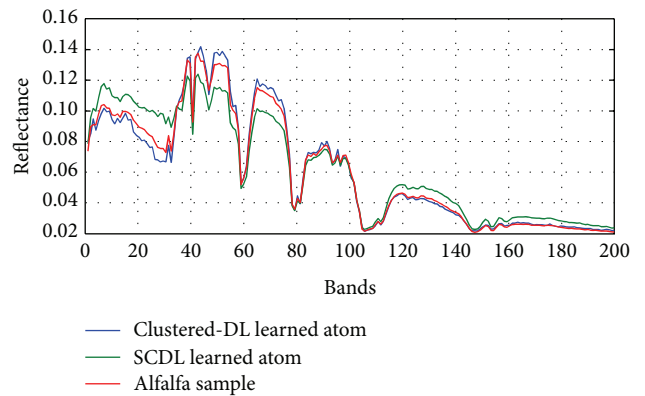


FIGURE 3: The comparison map of sample spectra for Alfalfa in the Indian Pines dataset and the learned dictionary atom obtained by Cluster-DL and SCDL.

spectra. The two obvious gaps in the spectra correspond to the regions of water absorption, which were removed.

4.2. The 2nd Experiment. In this experiment, we chose to focus on the Center of Pavia (shown in Figure 4(a)); our samples were collected in 2003 by the ROSIS sensor with a spatial resolution of 1.3 m/pixel in 115 spectral bands covering $0.43 \mu\text{m}$ to $0.86 \mu\text{m}$. Figures 4(b) and 4(c) show the ground-truth and classification maps obtained by SVM, whereas

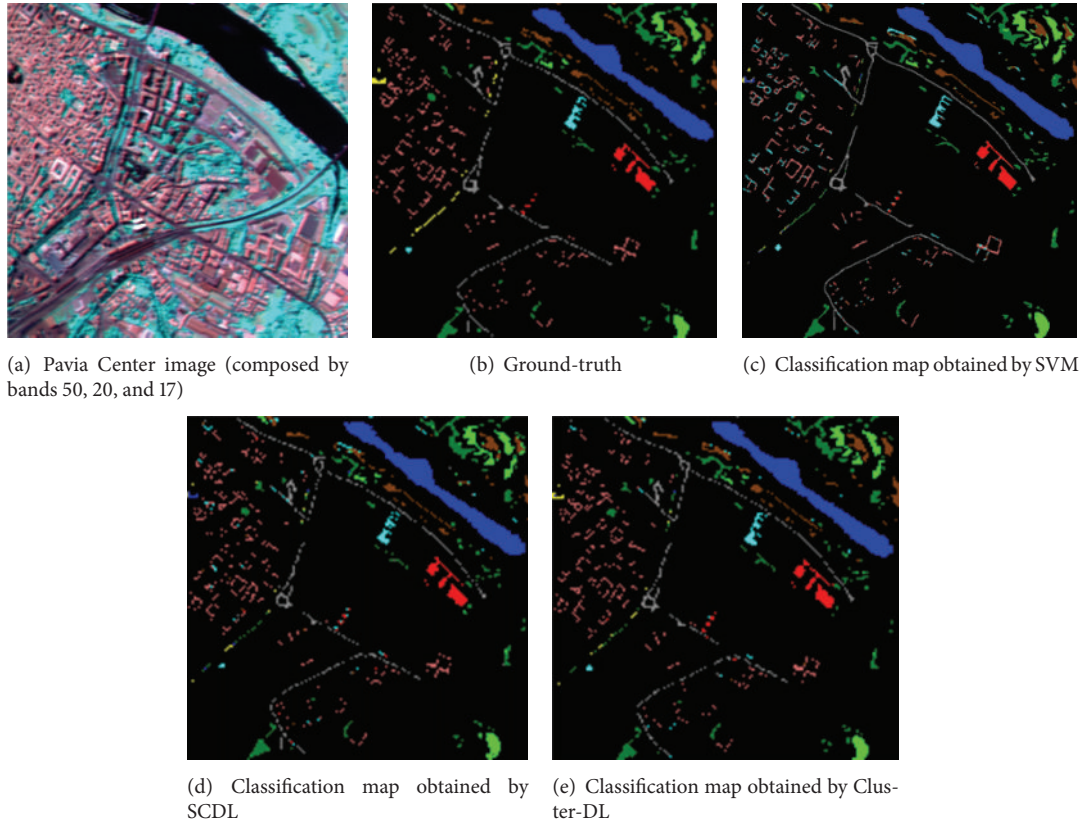


FIGURE 4: Pavia Center hyperspectral image and the comparison maps of different classification.

TABLE 1: Indian Pines ground-truth classes and train/test sets.

Number	Class	Train	Test
1	Alfalfa	6	48
2	Corn-notill	144	1290
3	Corn-min	84	750
4	Corn	24	210
5	Grass/pasture	50	447
6	Grass/trees	75	672
7	Grass/pasture-mowed	3	23
8	Hay-windrowed	49	440
9	Oats	2	18
10	Soybeans-notill	97	871
11	Soybeans-min	247	2221
12	Soybean-clean	62	552
13	Wheat	22	190
14	Woods	130	114
15	Building-grass-trees-drives	38	342
16	Stone-steel towers	10	85

Figures 4(d) and 4(e) show the classification maps obtained by SCDL and Cluster-DL (Table 3).

According to the experimental results, the clustered dictionary learning algorithm proposed in this paper can significantly improve classification accuracy. In the 1st experiment,

TABLE 2: Classification accuracy and execution time (S) for AVIRIS Indian Pines for different classifiers.

Classifiers	SVM	SCDL	Cluster-DL
OA	0.7678	0.9664	0.9679
AA	0.6752	0.9401	0.9434
Kappa	0.7301	0.9617	0.9634
Time (S)	1.9765	46.204	19.4878

TABLE 3: Classification accuracy for Pavia Center for different classifiers.

Classifiers	SVM	SCDL	Cluster-DL
OA	0.9122	0.9488	0.9734
AA	0.8491	0.8970	0.9320
Kappa	0.8941	0.9378	0.9678
Time (S)	0.50	68.34	4.73

Cluster-DL can improve classification accuracy from 0.9664, without the clustered dictionary learning algorithm, to 0.9679. In the 2nd experiment, Cluster-DL can improve the classification accuracy from 0.9488 to 0.9734; this means that the clustered dictionary learning algorithm has more obvious advantages when the terrain is more complex. The execution time of the SVM algorithm is less than the time of the dictionary learning algorithm, which also illustrates the fact

that the clustered structural dictionary learning improved the classification accuracy by increasing the execution time.

5. Conclusion and Discussion

In this paper, we have investigated clustered dictionary learning algorithms based on the models of hyperspectral data for HSI classification. Our research represents a hyperspectral sample with a linear combination of a few atoms learned from the data. The identical clustered groups share the atoms of a dictionary. The hyperspectral samples are classified by a linear SVM trained on the coefficients of this linear combination. Experiments on two sets of real HSI data confirmed this model's effectiveness for HSI classification and show that the proposed method can achieve better overall accuracy and Kappa coefficients. This is because the basic SVM classification does not take into account the relationship between the pixels. The SCDL classification partitioned the pixels into groups of the same size; the features of the group are not necessarily similar when grouped by size. In this paper, hyperspectral image is partitioned into a number of different groups by k -means++; the pixels in the same cluster groups have revealed hidden relationships within the spectral bands; this is closer to real object. Further research is needed in order to better understand how to integrate information between spatial and spectral information of HSI and utilize supervised classification algorithms to improve the classification accuracy and execution time.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors were sponsored by the National Natural Science Funds (nos. 41372340 and 41071265). Sincere thanks are due to the Committee of Development Foundation of Sciences and Technology for Geology and Minerals, Ministry of Land and Resources (MLR), China, that provided the financial support for doing an advanced research of the project. Sincere thanks are due to Soltani-Farani A and Paolo Gamba for giving one of the authors a very friendly help.

References

- [1] G. Shaw and D. Manolakis, "Signal processing for hyperspectral image exploitation," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 12–16, 2002.
- [2] A. Plaza, J. A. Benediktsson, J. W. Boardman et al., "Recent advances in techniques for hyperspectral image processing," *Remote Sensing of Environment*, vol. 113, supplement 1, pp. S110–S122, 2009.
- [3] R. K. Robinson and S. A. Jennings, "Hyperspectral imaging on the international space station: an innovative approach to commercial development of space," in *Proceedings of the 42nd AIAA Aerospace Sciences Meeting and Exhibit*, pp. 10584–10591, Reno, Nev, USA, January 2004.
- [4] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [5] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.
- [6] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, New York, NY, USA, 2004.
- [7] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 6, pp. 1351–1362, 2005.
- [8] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [9] M. D. Iordache, J. Bioucas-Dias, and A. Plaza, "Dictionary pruning in sparse unmixing of hyperspectral data," in *Proceedings of the 4th IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pp. 1–4, Shanghai, China, 2012.
- [10] A. S. Charles, B. A. Olshausen, and C. J. Rozell, "Learning sparse codes for hyperspectral imagery," *IEEE Journal on Selected Topics in Signal Processing*, vol. 5, no. 5, pp. 963–978, 2011.
- [11] A. Soltani-Farani, H. R. Rabiee, and S. A. Hosseini, "Spatial-aware dictionary learning for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 1, pp. 527–541, 2014.
- [12] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B: Methodological*, vol. 58, pp. 267–288, 1996.
- [13] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *Journal of Machine Learning Research*, vol. 11, no. 1, pp. 19–60, 2010.
- [14] X.-F. Song and L.-C. Jiao, "Classification of hyperspectral remote sensing image based on sparse representation and spectral," *Journal of Electronics & Information Technology*, vol. 34, no. 2, pp. 268–272, 2012.
- [15] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2477–2488, 2005.
- [16] D. Arthur and S. Vassilvitskii, "K-means++: the advantages of careful seeding," in *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms*, Society for Industrial and Applied Mathematics, 2007.
- [17] C.-C. Chang and C.-J. Lin, "LIBSVM: a Library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, article 27, 2011.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

