

Hyperspectral Image Classification Using Discriminative Dictionary Learning

Y Zongze, S Hao, J Kefeng, Z Huanxin

College of Electronic Science and Engineering, National University of Defense Technology, Changsha, Hunan, P. R. China, 410073.

E-mail: yiyepianfeng@126.com

Abstract. The hyperspectral image (HSI) processing community has witnessed a surge of papers focusing on the utilization of sparse prior for effective HSI classification. In sparse representation based HSI classification, there are two phases: sparse coding with an over-complete dictionary and classification. In this paper, we first apply a novel fisher discriminative dictionary learning method, which capture the relative difference in different classes. The competitive selection strategy ensures that atoms in the resulting over-complete dictionary are the most discriminative. Secondly, motivated by the assumption that spatially adjacent samples are statistically related and even belong to the same materials (same class), we propose a majority voting scheme incorporating contextual information to predict the category label. Experiment results show that the proposed method can effectively strengthen relative discrimination of the constructed dictionary, and incorporating with the majority voting scheme achieve generally an improved prediction performance.

1. Introduction

Hyperspectral imaging sensors acquire digital images in hundreds of continuous narrow spectral bands spanning the visible to infrared spectrum [1]. Different materials usually are spectrally separable as they reflect electromagnetic energy differently at specific wavelengths. This discriminative ability based on their spectral characteristics contributes for HSI classification tasks.

Recently, inspired by the success of compressive sensing (CS), sparse representation model based classification (SRC) framework has been proposed to solve many computer vision and pattern recognition tasks. The query signal is collaboratively coded over a dictionary of atoms with some sparsity constraint, and then classification is performed based on the coding coefficients and the dictionary. Sparse model based algorithms have been investigated successfully in areas like signal reconstruction [2], motion segmentation [3], image super-resolution [4], face recognition [5], target detection and classification in remote sensing field [6], where the usage of sparsity as a prior often leads to state-of-the-art performance.

In sparse representation based HSI classification, there are two phases: sparse coding over a dictionary and classification. In this paper, the proposed optimization dictionary construction employs a similarity measure to capture the relative difference among different classes. The samples for a specific class are then selected according to the similarity measurement. The competitive selection strategy ensures that atoms in the resulting over-complete dictionary are the most discriminative. On the other hand, motivated by the assumption that spatially adjacent samples are statistically related and even belong to the same materials, a majority voting scheme incorporating contextual information is



introduced to predict the category label. Three groups of experiments are conducted on two public available dataset to verify the effectiveness of the proposed algorithm. The results demonstrated the efficiency and enhanced accuracy of the proposed methods.

The remainder of this paper is structured as follows. Section 2 provides a brief review of the sparse representation model based classification. Section 3 proposes the fisher discriminative dictionary learning method and the majority voting scheme. Section 4 presents the experiment results and comparisons. Finally, the paper is concluded in section 5.

2. Sparse representation model based HSI classification

Naturally, the signals tend to have a representation biased towards their own class, i.e. the sparse representation is mainly formed from its own class. The dictionary consists of randomly chosen samples of each specific class. The final label of the query pixel is assigned with the sparse coefficients which encodes the class discriminative information.

Suppose $\mathbf{x} \in \mathbb{R}^{M \times 1}$ is an M-dimensional unlabelled hyperspectral data. Then \mathbf{x} can be written as a sparse linear combination of all the training samples as

$$\mathbf{x} = \mathbf{A}\boldsymbol{\theta} \quad (1)$$

where \mathbf{A} is a $M \times N$ dictionary whose each column corresponds to training samples of every class. $\boldsymbol{\theta}$ is an unknown sparse vector. Given the dictionary \mathbf{A} , we can formulate the sparse representation in the following equation

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \|\boldsymbol{\theta}\|_0 \quad s.t. \quad \mathbf{x} = \mathbf{A}\boldsymbol{\theta} \quad (2)$$

where the $\|\cdot\|_0$ denotes the L_0 norm which is defined as the number of the nonzero entries in the vector. The aforementioned problem is NP-hard, fortunately they can be approximately solved with three major methods. If $\boldsymbol{\theta}$ is sparse enough, the L_0 constraint can be relaxed to L_1 norm, then convex programming techniques [7] can be adopted to solve the problem. Another solution is greedy pursuit algorithms such as Matching Pursuit (MP) [8], Orthogonal Matching Pursuit (OMP) [9] algorithm, Simultaneous Orthogonal Matching Pursuit (SOMP) algorithm [10] and Subspace Pursuit (SP) [11]. And also the iterative thresholding [12, 13] is effective. In this paper we use the L_1 toolbox to pursue the sparse vector.

When the sparse vector is obtained, the class of observation \mathbf{x} can be determined directly by the characteristics of the recovered sparse vector $\hat{\boldsymbol{\theta}}$. It assigns the class with the minimum residual to \mathbf{x}

$$\mathbf{x} \in \text{class}(K) \quad \text{when} \quad r^K(\mathbf{x}) = \min_{1 \leq K \leq C} r^K(\mathbf{x}) \quad (3)$$

where the K th residual is defined as $r^K(\mathbf{x}) = \|\mathbf{x} - \mathbf{A}^K \boldsymbol{\theta}^K\|_2$ $K=1,2,\dots,C$. And \mathbf{A}^K denotes the subdictionary and $\boldsymbol{\theta}^K$ denotes the coefficients corresponding to \mathbf{A}^K .

3. Discriminative Dictionary Learning and Majority Voting Scheme

3.1. Fisher discriminative dictionary learning

In sparse representation model based hyperspectral image classification, the dictionary usually must be predefined. Chen Yi *et al.* [14,15] randomly selected the labelled training samples of all classes as the dictionary to code the query hyperspectral data. However the lazy dictionary used in [14,15] may not be effective enough to represent the query image due to the uncertain and noisy information in the original training samples. And also randomly chosen training samples can be not fully representing the discriminative information of each class. Meng Yang *et al.* [16] proposed fisher discriminative dictionary learning (FDDL), which aims to learn the space where the given signal could be well represented or coded for processing from the training samples. One contribution of this paper is introducing the FDDL method in hyperspectral image to achieve discriminative and concise or fuzzy dictionary.

According to aforementioned description, a structured dictionary $\mathbf{D}=[\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_C]$ needs to be learned, where \mathbf{D}_i is the class-specified sub-dictionary associated with the class A_i , and C is the total number of all classes in hyperspectral image of interest. The set of training samples is denoted by $\mathbf{A}=[\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_C]$, where \mathbf{A}_i represents the subset of the training samples of class i . The coding coefficients of \mathbf{A} over \mathbf{D} is denoted by \mathbf{X} , i.e. $\mathbf{A} \approx \mathbf{DX}$, where $\mathbf{X}=[\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_C]$ and \mathbf{X}_i is the sub-matrix containing the coding coefficients of \mathbf{D}_i over \mathbf{A}_i . Apart from requiring that \mathbf{D} should have powerful reconstruction capability of \mathbf{A} , we also require that \mathbf{D} should have powerful discriminative capability of images in \mathbf{A} . Finally the FDDL model can be presented as:

$$J_{(\mathbf{D}, \mathbf{X})} = \arg \min_{(\mathbf{D}, \mathbf{X})} \left\{ \sum_{i=1}^C r(\mathbf{A}_i, \mathbf{D}, \mathbf{X}_i) + \lambda_1 \|\mathbf{X}\|_1 + \right. \\ \left. (tr(S_w(\mathbf{X}) - S_B(\mathbf{X})) + \eta \|\mathbf{X}\|_F^2) \right\} \quad (4)$$

where $r(\mathbf{A}, \mathbf{D}, \mathbf{X})$ is the discriminative fidelity term and is defined as

$$r(\mathbf{A}_i, \mathbf{D}, \mathbf{X}_i) = \|\mathbf{A}_i - \mathbf{DX}_i\|_F^2 + \|\mathbf{A}_i - \mathbf{D}_i \mathbf{X}_i^i\|_F^2 + \sum_{j \neq i}^C \|\mathbf{D}_j \mathbf{X}_i^j\|_F^2 \quad (5)$$

where $\|\mathbf{A}_i - \mathbf{DX}_i\|_F^2$ denote the representation of \mathbf{A}_i over \mathbf{D} . Since \mathbf{A}_i should be well represented by \mathbf{D}_i but not by $\mathbf{D}_j, j \neq i$. That implies that corresponding coefficients \mathbf{X}_i^i should $\|\mathbf{A}_i - \mathbf{DX}_i^i\|_F^2$ be small, while the term \mathbf{X}_i^j should have nearly zero coefficients such that $\|\mathbf{D}_j \mathbf{X}_i^j\|_F^2$ is small. $\|\mathbf{X}\|_1$ is the sparsity constraint; To make the coding coefficients discrimination, i.e. \mathbf{X} , according to Fisher discrimination criterion[17], by minimizing the within-class scatter of \mathbf{X} , denoted by $S_w(\mathbf{X})$, and max the between-class scatter of \mathbf{X} , denoted by $S_B(\mathbf{X})$. Because the function $tr(S_w(\mathbf{X}) - S_B(\mathbf{X}))$ is non-convex and unstable, the elastic term $\eta \|\mathbf{X}\|_F^2$ was added to overcome the problem.

3.2. Majority voting scheme

Apart from the spectral signature, contextual information is proved important for HSI classification. In empirical observations spatially adjacent samples are statistically related and even belong to the same materials (same class). Hyperspectral image classification based on sparse representation usually is performed on spectral information of single pixel independently. However, the spatial/contextual information lacks enough attention. In this paper, In addition to the constraints on sparsity and reconstruction accuracy when recovering the sparse vector, we exploit the smoothness of the HSI pixels, which indicates that the neighboring HSI pixels have similar spectral characteristics as well as belong to the same class [15]. We propose a majority voting scheme, which incorporates contextual information to predict the category label. The class of test sample is now assigned by the local neighboring pixels instead of itself. We define the 4 neighbors' constraint as Laplacian smoothness constraint (LSC) and 8 neighbors' constraint as Auto-Regressive smoothness constraint (ARSC). In proposed voting scheme, the class label of center pixel x is voted by the local neighboring (4 or 8) pixels and the most ballots win, i.e. if half of the pixels containing x and its spatial neighbors are belonging to K , then the x is labeled as K .

4. Experiments and analysis

The test hyperspectral image in our experiments is the 220-bands Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) image Indiana Pine of spatial dimension 145×145 . It contains 16 ground-truth classes, most of which are different kinds of crops. For each class, we randomly choose 10% of the labelled samples for training samples and the rests for testing, which is presented in Figure 1.

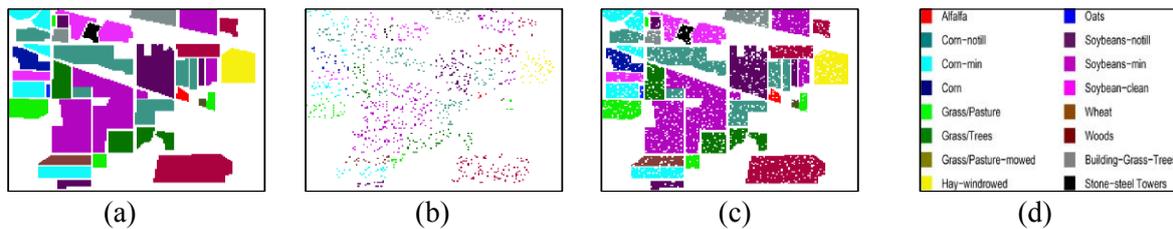


Figure 1. AVIRIS Indian Pine Data Set: (a) Ground Truth; (b) Training samples; (c) Test samples; (d) Class information.

Firstly, we verify the performance of FDDL on binary separable problem. *Corn_notill* and *Corn_min* are experimented. We compare the classification results obtained by FDDL dictionary and the one of the original dictionary. The number of atoms and classification accuracy (CA) is evaluated.

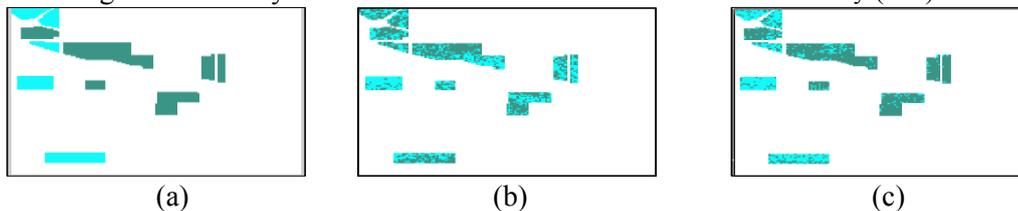


Figure 2. *Corn_notill* and *Corn_min* samples: (a) Ground Truth; (b) Classification Result with Usual Dictionary; (c) Classification Result with FDDL Dictionary.

Table 1. Comparison between usual dictionary and FDDL Dictionary.

Item	Atoms	Classification Accuracy (CA)	
		<i>Corn_notill</i>	<i>Corn_min</i>
Usual Dictionary	180	63.95%	62.58%
FDDL Dictionary	151(Average)	89.32%	86.71%

Then, two experiments are conducted to demonstrate the advantage of the majority voting scheme. In group one, the *Soybean_notill* and *Soybean_min* are picked out. And in group two, the *Corn_notill* and *Corn_min* are selected. Two kinds of smoothness constraints are applied to improve the original classification result. Three maps obtained by the Laplacian smoothness constraint, AR smoothness constraint and these two constraints simultaneously are observed. Classification performances are finally compared visually and quantitatively to the original one. The average classification accuracy is illustrated as below.

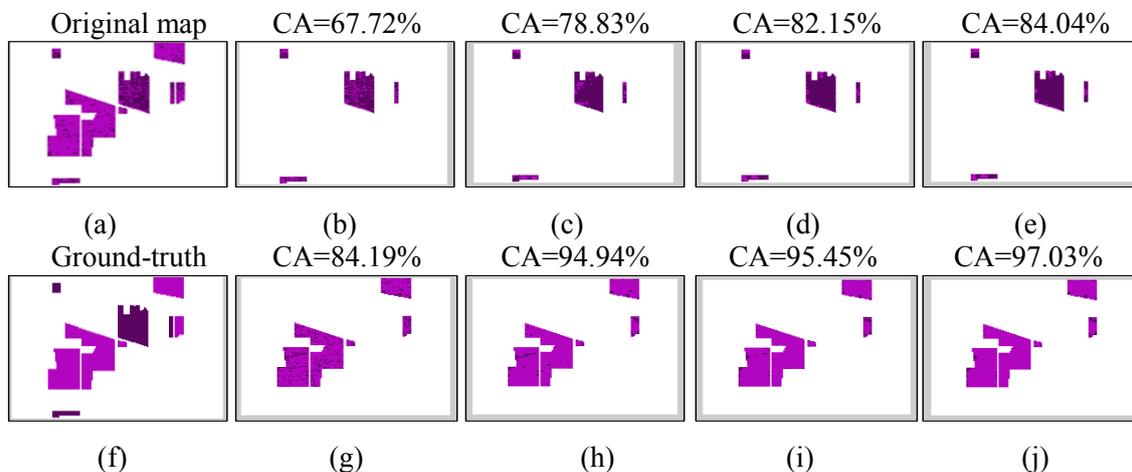


Figure 3. Classification maps and classification accuracy (CA) with smoothness constraint: (a) Original map; (b) Original classification map of *Soybean_notill*; (c) LSC; (d) ARSC; (e) LSC and ARSC; (f) Ground-truth map; (g) Original classification map of *Soybean_min*; (h) LSC; (i) ARSC; (j) LSC and ARSC.

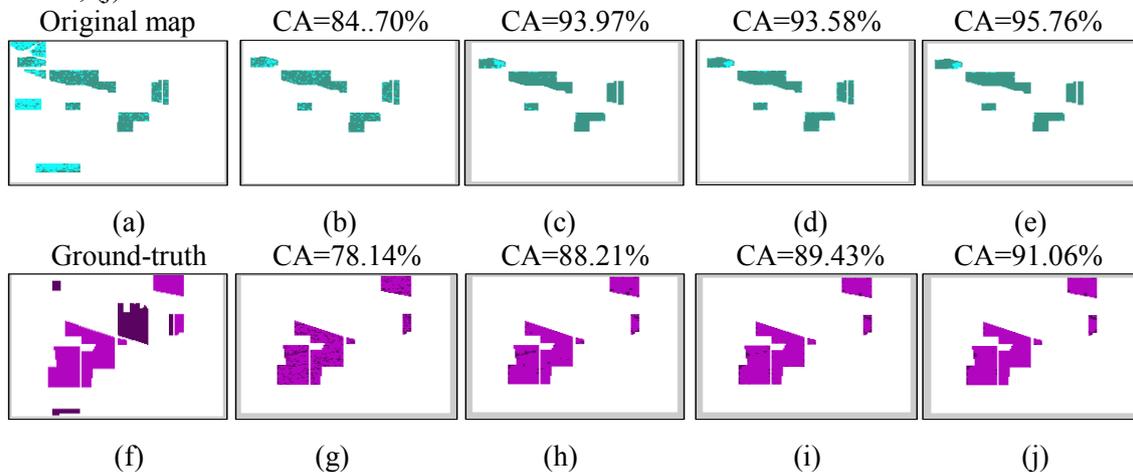


Figure 4. Classification maps and classification accuracy (CA) with smoothness constraint: (a) Original map; (b) Original classification map of *Corn_notill*; (c) LSC; (d) ARSC; (e) LSC and ARSC; (f) Ground-truth map; (g) Original classification map of *Corn_min*; (h) LSC; (i) ARSC; (j) LSC and ARSC.

We also illustrate the effect on the whole image classification performance. Figure 5 (c) shows the improved classification map with two constraints simultaneously. Then we apply FDDL and majority voting scheme simultaneously to the binary separable problem, We directly use the majority voting scheme on the basis of experiment result shown in Figure 5. We can see classification accuracy is improved significantly. More detail can be seen in Figure 6 and Table 2.

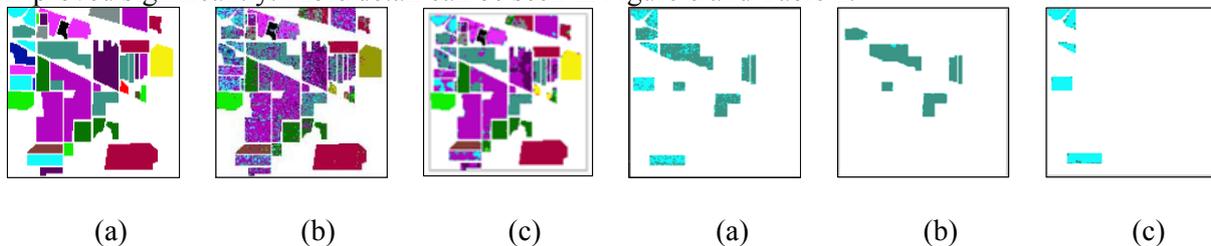


Figure 5. Classification maps and overall accuracy (OA): (a) Ground-truth map; (b) Original classification map (OA= 64.6838%); (c) Improved classification map (OA=87.665%);

Figure 6. The effect of FDDL and majority voting scheme: (a) Classification map obtained with FDDL dictionary; (b) Classification map of *Corn_notill*; (b) Classification map of *Corn_min*.

Table 2. Classification rate with majority voting scheme.

Item	Atoms	Original		Classification Accuracy with Voting Scheme	
		<i>Corn_notill</i>	<i>Corn_min</i>	<i>Corn_notill</i>	<i>Corn_min</i>
Usual Dictionary	180	63.95%	62.58%	85.34%	84.65%
FDDL Dictionary	151	89.32%	86.71%	97.14%	92.45%

5. Conclusion

In this paper, We first apply a novel fisher discriminative dictionary learning (FDDL) method to improve the dictionary in the sparse representation model. The FDDL aims to learn a structured dictionary whose sub-dictionaries have specific class labels. The FDDL's discrimination ability is that each sub-dictionary of the whole learned dictionary has good representation power to the samples of the corresponding class. Then, motivated by the assumption that spatially adjacent samples are statistically related and even belong to the same materials (same class), we present a novel majority voting scheme, which use reconstruction error and contextual smoothness character simultaneously. Experiment results show that FDDL can efficiently achieve a more discriminative and concise dictionary. And the proposed majority voting scheme utilizing spatial information incorporating with spectral information significantly improves the classification performance. Extensive experiments show our approach achieve generally a better prediction performance.

References

- [1] Borengasser M, Hungate W S and R. Watkins 2008 Hyperspectral Remote Sensing-Principles and Applications *Boca Raton, FL, USA: CRCPress*.
- [2] Candes E J and Tao T 2006 Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Transactions on Information Theory*, **52**:5406 – 5425.
- [3] Rao S, Tron R, Vidal R and Ma Y 2008 Motion segmentation via robust subspace separation in the presence of outlying, incomplete, or corrupted trajectories *CVPR*, 1–8.
- [4] Elhamifar E and Vidal R 2009 Sparse subspace clustering *CVPR*, 2790–2797.
- [5] Naseem I, Togneri R and Bennamoun M 2009 Sparse Representation for Video-Based Face Recognition, *Berlin, Germany: Springer-Verlag* 219–228.
- [6] Umamahesh S, Chen Y, Vishal M, Nasser M N and Trac D T 2013 Exploiting Sparsity in Hyperspectral Image Classification via Graphical Models. *IGRSS* **10**:505-509.
- [7] Tropp J A and Wright S J 2010 Computational methods for sparse solution of linear inverse problems *Proceedings of the IEEE* **98**: 948–958.
- [8] Shaykh A, Miloslavsky O K, Eugene, Nomura and Toshio 1990 Video compression using matching pursuits *IEEE T CIRC SYST VID* **9**: 123-143.
- [9] Cotter S F and Rao B D 2002 Sparse channel estimation via matching pursuit with application to equalization, *Communications IEEE Transactions on* **50**:374-377.
- [10] Tropp J A and Gilbert A C 2007 Signal recovery from random measurements via orthogonal matching pursuit, *Information Theory, IEEE Transactions on* **53**: 4655-4666.
- [11] Dai W and Milenkovic O 2009 Subspace pursuit for compressive sensing signal reconstruction, *Information Theory, IEEE Transactions on* **55**: 2230–2249.
- [12] Bobin J, Starck J and Ottensamer R 2008 Compressed Sensing in Astronomy, *IEEE J. Sel. Topics Signal Process* **2**: 718–726.
- [13] Daubechies I, Defrise M and De M C 2004 An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Common. Pure Appl. Math*, **57**:1413–1457.
- [14] Chen Y, Nasser M N and Trac D T 2010 Classification for Hyperspectral Imagery Based on Sparse Representation *WHISPERS*.
- [15] Chen Y and Nasser M N 2011 Hyperspectral Image Classification via Kernel Sparse Representation *ICIP*, 1233-1236.
- [16] Meng Y, Zhang L, Feng X C and David Z 2011 Fisher Discrimination Dictionary Learning for Sparse Representation, *ICCV* 543-550
- [17] Duda R, Hart P and Stork D 2000 Pattern classification (2nd ed.), *Wiley-Interscience*.