CLASSIFICTION FOR HYPERSPECTRAL IMAGERY BASED ON NONLOCAL WEIGHTED JOINT SPARSITY MODEL

Jiayi Li[1], *Hongyan Zhang*[1], *Yuancheng Huang*[2], *and Liangpei Zhang*[1]

1: The State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, P.R. China

2: The Center for Optical Imagery Analysis and Learning, State Key Laboratory of Transient Optics and Photonics, Xi'an, P.R. China

e-mail:zhanghongyan@whu.edu.cn

ABSTRACT

A nonlocal weighted joint sparse representation classification method for hyperspectral image is proposed in this paper. A discriminated contributions based on nonlocal spatial structure information are utilized in the joint sparsity model framework. The proposed algorithm is tested on two hyperspectral images. Experimental results suggest that the proposed algorithm shows superior performance over other sparsity-based algorithms and the classical hyperspectral classifier SVM.

Index Terms—nonlocal, joint sparse representation, hyperspectral image classification

1. INTRODUCTION

Hyperspectral imagery (HSI) which consists in observing the same scene at different wavelengths refers to that every pixel of the image is represented by hundreds of values, each corresponding to a different narrow wavelength [1]. Image classification, which aims at categorizing all pixels in a remote sensing image into one of several land cover classes, is an important application of HSI. Up to date, a lot of HSI classification techniques [2] have been proposed, such as SVM [3], decision trees, artificial neural networks and *etc*.

In recent years, as a popular signal modeling technique, sparse representation has been widely used in image processing and analysis. For HSI classification, Chen *et. al.* proposed a classification method based on joint sparsity model (JSM)[4] that obtained excellent classification results [5]. Inspired by their work, we propose a nonlocal weighted joint sparse representation classification (NLW-JSRC) method for HSI in this paper. The weighted joint sparsity model (W-JSM) is built up with hyperspectral pixels in a small neighborhood around the central test pixel. The weight of one specific neighborhood pixel is measured by the similarity between the neighborhood pixel and the central test pixel, which is often referred as nonlocal weighted scheme. After all the weights of the pixels in the neighborhood are determined, we set up the nonlocal weighted joint sparsity model which is solved by the simultaneous orthogonal matching pursuit (SOMP) [6] algorithm.

The remaining part of this paper is organized as follows. Details about the proposed NLW-JSRC algorithm are described in Section 2. Section 3 shows the experimental results conducted on several HSIs with the proposed algorithm and several state-of-the-art classification methods. Finally, Section 4 summarizes our work.

2. CLASSIFICATION OF HSI USING SPARSE REPRESENTION

In this section, we introduce the nonlocal weighted joint sparse representation classification algorithm which discriminates the contribution of one specific neighborhood pixel to the central test pixel by the similarity of the two image patches centered at the corresponding neighborhood pixel and the central test pixel, respectively.

2.1. Sparse representation

In the sparsity model, it is assumed that a signal can be approximated by a sparse liner combination of elements from a basis set named overcompleted dictionary. We consider a $B \times N$ dictionary matrix A with $B \ll N$, the compact signal $s \in \mathbb{R}^B$ can be approximately represented by multiplying the dictionary with a sparse vector α in which only a few elements are nonzero. The

978-1-4799-3406-5/12/$31.00©2012IEEE

sparse vector can be represented by solving the following optimization problem:

$$\alpha = \mathbf{arg\,min}\|\alpha\|_0 \quad s.t.\ \ s = A\alpha + \xi \tag{1}$$

where ξ is a small constant for noise of the signal.

2.2. Joint sparse representation classification (JSRC)

For sparse representation classification (SRC) [5], suppose we have M distinct classes and stack the given $N_i\ (i = 1,\dots,M)$ training pixels as columns of a sub dictionary $A_i = [a_{i,1}, a_{i,2}, \dots a_{i,N_i}] \in \mathbb{R}^{B\times N_i}$, denote A_i as a linear low-dimensional space, and B as the band number of the HSI. The signal s which belongs to i th class can be compactly represented as a linear combination of the given training, *i.e.*, $s \approx A_i\alpha_i$. As the identity of the signal s is initially unknown before classification, in order to linearly represent it, we define a new matrix $A \in \mathbb{R}^{B\times N}$ which include all the training pixels and consider the A_i as a sub matrix of A. Signal s can be represented as:

$$\begin{aligned} s &= A_1\alpha_1 + \dots + A_i\alpha_i + \dots + A_M\alpha_M + \varepsilon \\ &= \underbrace{[A_1 \dots A_i \dots A_M]}_{A}\underbrace{[\alpha_1 \dots \alpha_i \dots \alpha_M]^T}_{\alpha} + \varepsilon \\ &\approx A\alpha \in \mathbb{R}^B \end{aligned} \tag{2}$$

As the signals from the same class often span in a same low-dimensional subspace, which is constructed by the active training samples corresponding to the nonzero entries of the sparse vector α, we classify the signal s based on the approximations by assigning it to the object class that minimizes the residual:

$$\begin{aligned} class(s) &= \mathbf{arg}\ \underset{i=1,\dots,M}{\mathbf{min}}\ r_i(s) \\ &= \mathbf{arg}\ \underset{i=1,\dots,M}{\mathbf{min}}\ \|s - A_i\alpha_i\|_2 \end{aligned} \tag{3}$$

For hyperspectral data, each pixel in a small neighborhood with similar spectrum can be linearly represented in a same low-dimensional feature subspace by different compact coefficients. Assuming that a single signal can be sparsely represented by the $B\times N$ structured dictionary A, let $S = [s_1, \dots, s_T]$ be a $B\times T$ signal matrix constructed with a patch of hyperspectral image whose central pixel is the test sample and all the columns of S meet the "common sparse supports", and then the joint signal matrix can be represented by JSM:

$$\begin{aligned} S &= [s_1\ s_2\ \dots\ s_T] = [A\alpha_1 + \varepsilon_1\ A\alpha_2 + \varepsilon_2 \dots A\alpha_T + \varepsilon_T] \\ &= A[\alpha_1\ \alpha_2 \dots \alpha_T] + \Sigma = A\underbrace{[\alpha_1\ \alpha_2 \dots \alpha_T]}_{\Psi} + \Sigma \\ &\approx A\Psi \end{aligned} \tag{4}$$

where Σ is the noise matrix which contains noise of each signal in the joint matrix. The optimization model corresponding to (4) can be expressed as:

$$\Psi = \mathbf{arg\,min}\|\Psi\|_{row,0} \quad s.t.\ \ S = A\Psi + \Sigma \tag{5}$$

where $\|\bullet\|_{row,0}$ denotes the numbers of nonzero rows of Ψ .Once the sparse matrix Ψ is obtained, following the similar way as SRC, we label the central test pixel of the patch by calculating the residuals [5]:

$$class(s_c) = \mathbf{arg}\ \underset{i=1,\dots,M}{\mathbf{min}}\ \|S - A_i\Psi^i\|_F \tag{6}$$

where Ψ^i denotes the portion of the recovered sparse coefficients corresponding to the training samples of the i th class.

2.3. Weighted joint sparse representation classification (W-JSRC)

Adjacent hyperspectral pixels often consist of similar materials, so the correlations of these spectral signals are high. However, it is not fair to take low-correlation pixels into JSM and consider them equally as the high-correlation pixels. When there exist the pixels in the neighboring window, which are close to the central pixel in spatial distribution, have significant disparities to the central pixel in spectral curve, the action that stacks these low-correlation pixels in the joint sparse representation classification method is unwise. Thus, the assumption of the joint sparsity model will not be satisfied.

In view of this, each pixel in the neighborhood window should have different weight to the classification of the central pixel, and the weight should be determined by the correlation between each neighborhood pixel and the central pixel. So we extend the JSRC to W-JSRC by incorporating weighting as follows:

$$\Psi_W = \mathbf{arg\,min}\|\Psi_W\|_{row,0} \quad s.t.\ \ SW = A\Psi_W + \xi \tag{7}$$

where $W = diag(w_1, w_2, \dots, w_T)$ is a diagonal matrix, and each entry on its main diagonal denotes the contribution of the corresponding neighborhood pixel to the central test one.

978-1-4799-3406-5/12/$31.00©2012IEEE

2.4. Nonlocal weighted joint sparse representation (NLW-JSRC)

In recent years, a nonlocal [8] weigh scheme which aims at explicitly exploiting self-similarities in natural image has been widely used in sparse signal processing [9]. Assuming that p denotes the central pixel to be tested and q denotes a pixel in the neighborhood of p. The weighted scheme can be mathematically expressed as:

$$w(p,q) = \exp\left(-\frac{\|J(p) - J(q)\|}{\rho} \right) \tag{8}$$

where $\|J(p) - J(q)\|$ denotes the similarity measure between p and q, $J(\bullet)$ denotes spatial structure operators that centers at the corresponding image position, ρ is the factor to adjust the decay of the exponential function. We apply the nonlocal weighted scheme to the W-JSRC method, thus the nonlocal spatial information are included in the NLW-JSRC model. The nonlocal weighted joint sparsity model can be represented as:

$$\begin{aligned} &\Psi_{NLW} = \arg\min \|SW_{NL} - A\Psi_{NLW}\|_F \ \ s.t. \|\Psi_{NLW}\|_{row,0} \le L \\ &W_{NL} = diag(w_1(p,q_1), w_2(p,q_2), \ldots, w_T(p,q_T)) \end{aligned} \tag{9}$$

where T is the size of neighborhood window and the NLW-JSRC model in (9) is solved by SOMP algorithm in our paper. We label the identity of the central test pixel s_c by minimizing the residual as before:

$$class(s_c) = \underset{i=1,\ldots,M}{\arg\min} \|SW_{NL} - A_i \Psi_{NLW}'\|_F \tag{10}$$

3. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we demonstrate the effectiveness of the proposed algorithm on two hyperspectral images. The classical classifier SVMs [2], JSRC [5] is used as benchmarks in this paper. In addition, the OMP [10] are used to solve the SRC problems in (1) and also included in the comparisons.

The first hyperspectral image in this paper was gathered by AVIRIS sensor over the Indian Pines test site in North-western Indiana and consists of pixels and 224 spectral reflectance bands in the wavelength range 0.4–2.5$\times$10^(-6) meters. We have also reduced the number of bands to 200 with 24 water absorption bands removed. The false color image can be visually shown in Figure 1(a). This image contains 16 ground truth classes which can be visually shown in Figure 1(b). In this experiment, we randomly sample 9% of the data in each class as the training samples and the remaining as the test samples, and the detailed information is shown in Table 1. The classification accuracy for each class using different classifiers is also shown in Table 1 and the classification maps are shown in Figure 1(c)-(f). The residual and optimal tolerance parameters for each greedy algorithm mentioned above are default. The optimal parameters for the NLW-JSRC are $L = 30$ and $T = 81$, where the corresponding optimal neighboring size for JSRC is $T = 25$. Besides, the parameters for SVM are obtained by 10 fold cross-validation. It is shown from Table 1 that by incorporating the nonlocal spatial information, the proposed algorithm outperforms the other classification algorithms.

The second hyperspectral image used in this paper is the 115-band ROSIS image Centre of Pavia of size 776×485, for which we use only 102 bands with the 13 water absorption bands removed. The false color image can be visually shown in Figure 2(a). There are 9 classes of interests as shown by the ground truth map in Figure 2(b), in this experiment, we randomly sample 1% of the data in each class as the training samples and the remaining as the test samples, and the detailed is shown in Table 2 which contains the classification accuracy for each class using various and the classification maps are shown in Figure 2(c)-(f). It is observed that we can draw the same conclusion with the first experiment.

4. CONCLUSIONS

In this paper, we propose a new hyperspectral image classification algorithm based on nonlocal weighted joint sparsity model, which exploit different contributions of the neighborhood pixels to the classification of the central test pixel by incorporating nonlocal spatial structural information to support improved classification capabilities. The extensive experimental results clearly suggest the proposed NLW-JSRC method can achieve competitive classification results. Our further work will focus on other more reasonable weighted schemes which can be used to further improve the classification performance.

5. REFERENCES

[1] E. Christophe, D. Léger and C. Mailhes, "Quality criteria benchmark for hyperspectral imagery," *Geoscience and Remote Sensing, IEEE Transactions on,* vol. 43, pp. 2103-2114, 2005.

[2] C.I. Chang, "Hyperspectral imaging: Techniques for spectral detection and classification", *Springer Us*, vol. 1, 2003.

[3] B.E. Boser, I.M. Guyon and V.N. Vapnik, "A training algorithm for optimal margin classifiers," *ACM New York, NY, USA*, pp. 144-152, 1992.

[4] D. Baron, M.B. Wakin, M.F. Duarte, S. Sarvotham and R.G. Baraniuk, "Distributed compressed sensing," ed, 2005.

[5] Y. Chen, N.M. Nasrabadi and T.D. Tran, "Hyperspectral Image Classification Using Dictionary-Based Sparse Representation," *Geoscience and Remote Sensing, IEEE Transactions on,* pp. 1-13, 2011.

[6] J.A. Tropp, A.C. Gilbert and M.J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Processing,* vol. 86, pp. 572-588, 2006.

[7] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry and Y. Ma, "Robust face recognition via sparse representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 31, pp. 210-227, 2009.

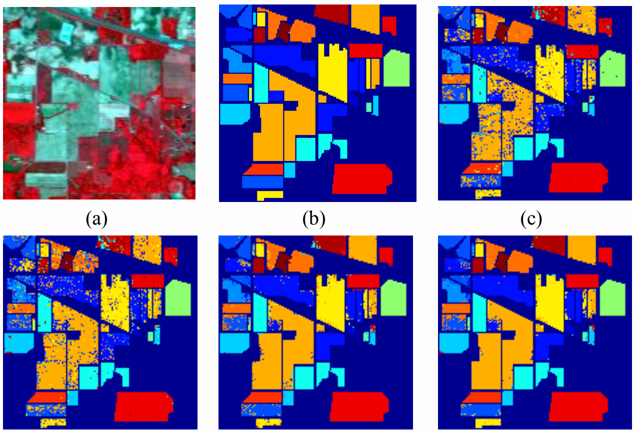
[8] A. Buades, B. Coll and J.M. Morel, "A non-local algorithm for image denoising," *Computer Vision and Pattern Recognition, CVPR 2005. IEEE Computer Society Conference on*, pp. 60-65 vol. 2, 2005.

[9] J. Mairal, F. Bach, J. Ponce, G. Sapiro and A. Zisserman, "Non-local sparse models for image restoration," *Computer Vision, 2009 IEEE 12th International Conference on*, pp. 2272-2279, 2009.

[10] J.A. Tropp and A.C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *Information Theory, IEEE Transactions on,* vol. 53, pp. 4655-4666, 2007.

6. ACKNOWLEDGEMENTS

This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2011CB707105, by the National Natural Science Foundation of China under Grants 40930532 and 61102112, by Post-doctoral Science Foundation of China under Grant 2011M501242.

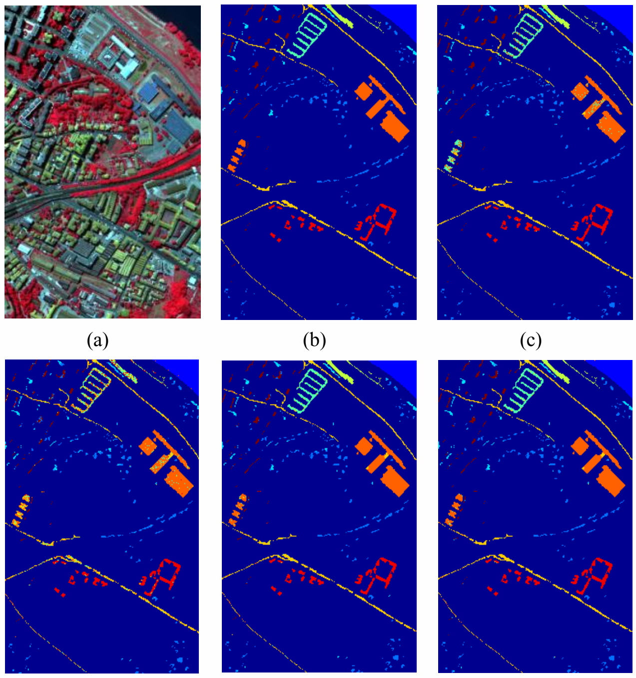


(a) (b) (c) (d) (e) (f)

Figure 1.Classification results of Indian Pines image: (a) false color image (R:57 G:27 B:17) (b) ground truth (c) SVM, (d) SRC, (e)JSRC, (f) NLW-JSRC

Table 1. Classification accuracy(%) for the Indiana Pine image on the test set using different classifiers

Class	Train	Test	SVM	SRC	JSRC	NLW-JSRC
1	6	40	0.5750	0.8000	0.9250	**0.9750**
2	129	1299	0.7282	0.7167	0.9145	**0.9169**
3	83	747	0.6988	0.7149	0.8527	**0.8862**
4	24	213	0.6291	0.5775	0.8122	**0.8873**
5	48	435	0.8874	0.8966	0.9172	**0.9448**
6	73	657	0.9802	0.9680	1	**1**
7	5	23	0.0867	**0.7826**	**0.7826**	**0.7826**
8	48	430	**1**	0.9977	**1**	**1**
9	4	16	0.3125	0.7500	0.3125	**0.8125**
10	97	875	0.6766	0.7166	**0.8766**	0.8686
11	196	2259	0.8309	0.7928	0.9464	**0.9695**
12	59	534	0.8783	0.6423	0.9195	**0.9457**
13	21	184	0.9511	0.9891	**1**	**1**
14	114	1151	0.9618	0.9513	0.9878	**0.9983**
15	39	347	0.6311	0.5994	0.8963	**0.9164**
16	12	81	0.8765	0.9506	**1**	0.9877
OA	958	9291	0.8182	0.7995	0.9313	**0.9467**
Kappa			0.7913	0.7708	0.9215	**0.9390**



(a) (b) (c) (d) (e) (f)

Figure 2.Classification results of Centre of Pavia image: (a) false color image, (R:102, G:56 ,B:31), (b) ground truth (c) SVM, (d) SRC, (e)JSRC, (f) NLW-JSRC

Table 2. Classification accuracy(%) for the Centre of Pavia image on the test set using different classifiers

Class	Train	Test	SVM	SRC	JSRC	NLW-JSRC
1	42	5268	**1**	0.9972	**1**	0.9994
2	35	3471	0.9015	0.8778	0.9231	**0.9268**
3	20	958	0.6378	0.8038	0.9509	**0.9541**
4	32	2108	0.7434	0.3439	**0.9132**	0.9032
5	22	1067	0.9156	0.5117	0.9897	**0.9963**
6	39	4850	0.8759	0.7781	**0.9462**	0.9454
7	58	7229	0.8434	0.8382	0.9644	**0.9772**
8	31	3091	0.8948	0.9673	0.9964	**0.9977**
9	24	1595	**1**	0.7643	0.9210	0.9442
OA	303	29637	0.8860	0.8227	0.9608	**0.9651**
Kappa			0.8655	0.7911	0.9538	**0.9588**

978-1-4799-3406-5/12/$31.00©2012IEEE