

Spatial Preprocessing for Improved Sparsity Based Hyperspectral Image Classification

Kavitha Balakrishnan, Sowmya V., Dr. K.P. Soman

Center for Excellence in Computational Engineering and Networking

Abstract

In this paper, we present that hyperspectral image classification based on sparse representation can be significantly improved by using an image enhancement step. Spatial enhancement allows further analysis of hyperspectral imagery, as it reduces the intensity variations within the image. Perona-Malik, a partial differential equation based non-linear diffusion scheme is used for the enhancement of the hyperspectral imagery prior to classification. The diffusion technique applied smoothens the homogenous areas of hyperspectral imagery and thereby increases the separability of the classes. The diffusion scheme is applied individually to each band of the hyperspectral imagery and it does not take into account the spectral relationship among different bands. Experiments are performed on the real hyperspectral dataset AVIRIS (Airborne Visible/IR Imaging Spectrometer) 1992 Indiana Indian Pines imagery. We compared the classification statistics of hyperspectral imagery before and after performing the spatial preprocessing step in order to prove the effectiveness of the proposed method. The experiments results proved that the hyperspectral image classification using sparse representation along with spatial enhancement step lead to 97.53% of classification accuracy which is high when compared with the classification accuracy obtained without applying the spatial preprocessing technique.

Keywords: *Compressed sensing, classification, sparse representation, hyperspectral imagery, Simultaneous OMP, preprocessing, Perona-Malik diffusion*

1. Introduction

Imaging spectrometry or hyperspectral sensing allows simultaneous acquisition of images in hundreds of narrow contiguous spectral bands spanning the visible to infrared region of the spectrum. The high resolution reflectance spectra collected by imaging spectrometers makes it possible to classify materials within the scene with improved accuracy. However,

hyperspectral imagery suffers from several drawbacks. The most important disadvantage is the computational complexity involved in processing high dimensional hyperspectral dataset. In addition to this, hyperspectral imagery suffers from noise and poor contrast due to different illumination conditions, degradation in transmission media, etc.

Various techniques have been proposed for sparsity based hyperspectral image classification. The sparsity model [2] [3], relies on the fact that hyperspectral pixels belonging to the same class will approximately lie in a low-dimensional subspace. Thus, a test pixel can be sparsely represented using few training samples from a dictionary and this sparse representation will give information about the class label. Since the hyperspectral images have large homogenous regions, pixels in the neighbourhood consist of similar material. Hence to exploit contextual information, a joint sparsity model [4] is employed where neighbouring pixels are represented by linear combination of few common training samples. Each pixel share the same training samples (atoms) but are weighted using different values. Previous approaches for hyperspectral image classification shows that use of kernel methods yields significant improvement in classification performance [6]. This is due to the fact that kernel methods project data into a nonlinear feature space making it more separable. Hence the linear sparse model is extended to a high dimensional feature space by using kernel function [7], where a test pixel is sparsely represented in terms of all training samples in a feature space by using a kernel based greedy algorithm. Further improvement in classification is brought about by using probabilistic graphical model [8], where discriminative tree graphs are learnt for each set of distinct sparse feature vectors. These features exhibit class conditional correlations and hence classification is performed by using a classifier that combines the distinct sparse features.

In this paper, we apply a spatial preprocessing step prior to sparsity based classification. The preprocessing technique performed in our proposed work is PDE based Perona-Malik nonlinear diffusion, which allows the homogenous regions to be smoothed out and thereby increases the separability between the classes.

However, the diffusion scheme is applied on each band individually and hence does not take into account the spectral relationship among different bands of the hyperspectral image. In order to prove the effectiveness of our proposed method, we have compared the classification statistics of the Indian Pines hyperspectral dataset before and after performing spatial preprocessing technique. The results have proved that hyperspectral image classification using sparse representation along with an enhancement step offers significantly better classification results.

2. Basic theory of Compressive Sensing

In this section, we present a brief overview of compressive sensing and sparse representation. Compressed sensing is a recently developing method that exploits the fact that signals are either sparse by itself or sparse in any other transformation domain. By the term sparsity, we mean that the information present in the signal can be compactly represented by using a few non-zero values. Consider a signal x of length N and with K non-zero entries, where $K \ll N$. Compressive sensing framework recovers the signal x from

$$y = Ax \quad (1)$$

where y is a vector and A is a dictionary matrix with size $M \times N$, where $M \ll N$. Here, we say that x is compressible in A . As $M \ll N$, gives infinitely many solutions for (1). One way of finding the solution for (1) is by finding the 'sparsest' x . The sparsest solution is obtained by solving the following optimization problem

$$\min \|x\|_0 \text{ subject to } Ax = y \quad (2)$$

where $\|x\|_0$ gives the count of the number of non-zero elements in x .

Hence, we find that sparse representation allows representing a signal as a linear combination of few atoms from the dictionary. The dictionary can be either based on a mathematical model of the data or it can be learned directly from the data.

3. Perona-Malik Non-Linear Diffusion Scheme

Spatial preprocessing techniques are used to remove noise and smoothen out the image. Hence, they enhance the quality of the image, resulting in improved classification accuracy. In this paper, PDE based diffusion schemes have been used as the preprocessing step.

In general, diffusion equation in 2-D is given by

$$\begin{aligned} \frac{\partial}{\partial t} u(x, y, t) &= \nabla \cdot (c(x, y, t) \nabla u(x, y, t)) \\ u(x, y, 0) &= u_0(x, y) \end{aligned} \quad (3)$$

where $c(x, y, t)$ is the diffusion conductance or diffusivity, ∇ is the gradient operator, $\nabla \cdot$ the divergence operator, and t denotes the number of iterations. If $c(x, y, t)$ is a constant, it leads to a linear diffusion equation. In that case, the filter not only smoothen out noise in the image, but also blurs edges as the image evolves in time. If the function $c(x, y, t)$ is not a constant and is image dependent, then the linear diffusion equation becomes a non-linear diffusion equation. By using a diffusion conductance c that was based on the derivative of the image at time t , we will be able to control the diffusion near the edges in the image. Such an improved formulation is known as anisotropic diffusion, where diffusion takes place according to a diffusion coefficient that is variable and adaptive in order to reduce the smoothing effect near edges. [11]

Perona-Malik [10] proposed anisotropic diffusion to smooth out noise and prevent diffusion at the edges. The term "anisotropic" as defined by Perona and Malik refers to the case where the diffusivity is a function varying with the location. However, the Perona and Malik process is essentially a nonhomogeneous nonlinear isotropic diffusion, since it utilizes a scalar-valued diffusivity function. The anisotropic diffusion equation given in (3) when applied to hyperspectral imagery is given as

$$\frac{\partial u_i}{\partial t} = \nabla \cdot (c \nabla u_i), i = 1, 2, \dots, B \quad (4)$$

where $u = (u_1, \dots, u_B)^T$ is the hyperspectral image composed of B bands.

The diffusivity function used here is

$$c(\|\nabla u\|) = \exp\left(\frac{-\|\nabla u\|^2}{k^2}\right) \quad (5)$$

where k is a constant which influences the smoothing process. A large value of k will cause low-contrast edge features to be smoothed out, while a small value of k leads to slow diffusion within homogeneous regions.

Now, $c(\|\nabla u\|)$ is a scalar function that controls the amount of diffusion. We find that $c(\|\nabla u\|) \rightarrow 0$ when $\|\nabla u\|$ is large. The norm of the image gradient $\|\nabla u\|$ is large at the edges of the image. This ensures that strong edges are less blurred by this function than noise and low-contrast details present in the image.

Hence, the basic idea in using the diffusivity function $c(\|\nabla u\|)$ is that it is variable and increase as

$\|\nabla u\|$ decreases. At homogenous regions in the image, where $\|\nabla u\|$ is small, $c(\|\nabla u\|)$ is large and lot of smoothing is done whereas in regions where $\|\nabla u\|$ is large (i.e. near edges), $c(\|\nabla u\|)$ tends to zero and as a result, the edges are preserved.

4. Classification using Sparse Representation with Contextual Information

4.1. Classification Algorithm

In this section, we brief out the sparsity based classification method using contextual information proposed by Yi Chen et al. [3]. Let $A_i = [a_{i,1}, a_{i,2}, \dots, a_{i,m}]$ be the training set corresponding to the i th class, where $a_{i,j}$, $j = 1, 2, \dots, n_i$ is a B -dimensional vector. A test sample $y \in \mathbb{R}^B$ from this class can be represented as a linear combination of the fewest training samples possible, out of the entire training set. Suppose we have M different classes, then $A = [A_1, A_2, \dots, A_M]$ be the concatenation of the N training samples from all M classes. Therefore, a test pixel can now be represented as a linear span of all training samples by $y = Ax$ where $x = [0, \dots, 0, \alpha_{i,1}, \dots, \alpha_{i,m}, 0, \dots, 0]^T$ is the sparse representation of y .

Now, in order to increase the classification accuracy, the spatial correlations among pixels are considered. This is known as the joint sparsity model [3]. Consider a neighbourhood N which consist of T pixels. Pixels are concatenated to form a matrix Y . Now, using joint sparsity model we can write

$$Y = [y_1 \ y_2 \ \dots \ y_T] = [Ax_1 \ Ax_2 \ \dots \ Ax_T] = AS \quad (6)$$

where $S = [x_1 \ x_2 \ \dots \ x_T]$. In joint sparsity model, hyperspectral pixels in a neighbourhood can be represented as linear combination of few common atoms from a dictionary. The matrix S can be recovered by solving the following joint sparse recovery problem

$$\hat{S} = \arg \min \|AS - Y\|_F \quad \text{subject to} \quad \|S\|_{\text{row},0} \leq K \quad (7)$$

where K denotes the number of non-zero elements. The simultaneous sparse recovery problems are NP-hard problems, which can be approximately solved by using greedy algorithm like Simultaneous Orthogonal Matching Pursuit (S-OMP) [5]. Finally, the residual for

each class is computed and the class for which residual is minimum is assumed to be the class of the test pixel.

4.2. Simultaneous Sparse Recovery Algorithm(S-OMP)

In simultaneous sparse approximation, several input signals are approximated using different linear combination of the same K elementary signals. This appears when we are analyzing multiple observations of a sparse signal that have been contaminated with noise.

Input: $B \times N$ dictionary matrix $A = [a_1, a_2, \dots, a_N]$, $B \times T$ matrix $Y = [y_1, y_2, \dots, y_T]$

- (1) Initialize the residual $\text{res} = Y$, the index set $\Lambda_0 = \emptyset$, and the iteration counter $k = 1$.
- (2) Find the index of the atom which best approximates all the T input signals

$$\lambda_k = \arg \max_{i=1, \dots, N} \|\text{res}_{k-1}^T \times a_i\|_p; p \geq 1$$

- (3) Update the index set

$$\Lambda_k = \Lambda_{k-1} \cup \lambda_k$$

- (4) Determine the orthogonal projector P_k onto the span of the atoms indexed in Λ_k .

$$P_k = (A_{\Lambda_k}^T A_{\Lambda_k})^{-1} A_{\Lambda_k}^T Y$$

- (5) Calculate the new residual

$$R_k = Y - A_{\Lambda_k} P_k$$

- (6) Increment k , and return to Step 2 until stopping criterion is met.

Output: Sparse representation \hat{S} whose non-zero rows indexed by index set Λ are the K rows of $P = (A_{\Lambda}^T A_{\Lambda})^{-1} A_{\Lambda}^T X$

Step 2 of the algorithm is referred to as the greedy selection. The intuition behind maximizing $\|\text{res}^T \times a_i\|_p$ is that we wish to find an atom that contributes the most energy to as many of the input signals as possible. In this paper we have selected $p = \infty$. It is important to note that each column of the residual is orthogonal to the atoms indexed in Λ . Therefore, no atom is ever chosen twice.

5. Proposed method

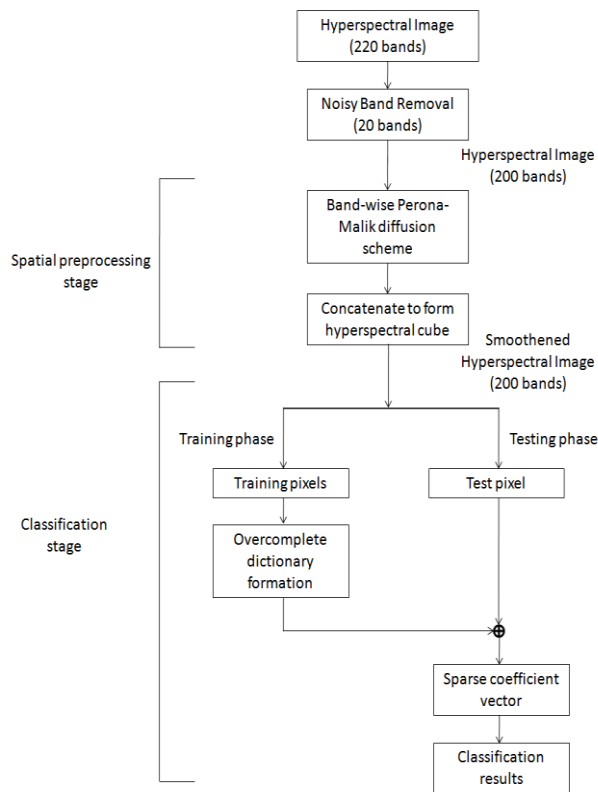


Figure 1. Block diagram of the proposed method

In literature, it can be seen that the most obvious way to improve classification performance is by using a preprocessing stage before the classification algorithm. In this paper, we have applied a Perona and Malik non-linear diffusion preprocessing step prior to sparsity based hyperspectral classification. Fig. 1 shows an overview of the proposed method. The input to the algorithm is the hyperspectral image Indian Pines, which consists of 220 spectral bands in the spectral range of 400-2500nm. The very next step is the removal of the noisy bands. Bands 104-108,150-163,220 are considered as noisy due to water absorption (Tadjudin and Landgrebe, 1998) and are hence removed. The resulting hyperspectral image consists of 200 spectral bands. The proposed algorithm mainly consists of two main stages-Preprocessing and Classification.

5.1. Spatial preprocessing step

In preprocessing step, Perona-Malik diffusion scheme is applied to each band of the hyperspectral

image. When performing band wise smoothing, contents of the other bands are ignored because each band of an image is treated “separately”. Hence, this mechanism does not take into account the spectral relationship among different bands of the hyperspectral image. Now, the diffused hyperspectral bands are combined to form the hyperspectral cube.

5.2. Classification step

Classification stage usually involves separating data into training and testing sets. In the training phase, 10% of the pixels of the hyperspectral image are randomly selected from each class and are concatenated to form a dictionary matrix A . Now, all the columns of the dictionary matrix must have unit norm. In testing phase, all the hyperspectral pixels, excluding background pixels are used for validation.

In a real hyperspectral image, neighbouring pixels are highly correlated mainly due to the fact that homogenous regions in the hyperspectral image are generally large when compared to the size of the pixel. Hence, in order to improve the classification accuracy information each pixel is classified based on its own spectral information along with the information extracted from its neighbourhood. This is what is done in the joint sparsity model, where pixels are selected for testing by using a window whose size is variable. The centre pixel of the window forms the test pixel, whose label is to be determined. The pixels in the neighbourhood are then arranged to form a matrix, where each column of the matrix represents a 200 dimensional vector. This matrix is represented as Y . This is illustrated below by using an example. Consider a hyperspectral image shown in Fig. 2. Here, we consider a 3×3 neighbourhood.

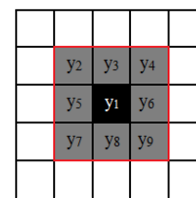


Figure 2. Hyperspectral Image

Let, y_1 be the test pixel and y_i , $i = 2, \dots, 9$ be the neighbouring pixels. Now, the pixel in the center is classified by taking into account the information from its neighbouring pixels. Each pixel is then vectorized and concatenated to form a matrix, which in this case is of size 200×9 .

Finally, Y is sparsely represented using dictionary matrix A . The matrix S can be recovered by solving the following joint sparse recovery problem

$$\hat{S} = \arg \min \|AS - Y\|_F \text{ subject to } \|S\|_{row,0} \leq K$$

where K denotes the number of non-zero elements. The above optimization problem is approximately solved by using greedy algorithm like Simultaneous Orthogonal Matching Pursuit (S-OMP) [5], which is explained in the previous section. The sparse matrix S , is a row sparse matrix since only a very few atoms of A is used to represent Y . The residual for each class is computed by calculating the difference between the original samples and its approximation.

$$r^m(Y) = \|Y - A^m \hat{S}^m\|_F, m = 1, 2, \dots, M$$

Finally, the class for which residual is minimum is assumed to be the class of the test pixel (center pixel).

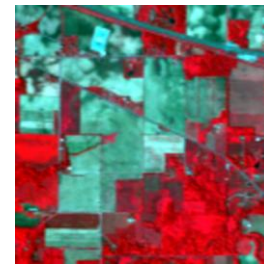
$$Class(y_i) = \arg \min_{m=1,2,\dots,M} r^m(Y)$$

6. Experimental Results

In this section, we present an overview of the characteristics of the data used followed by the accuracy assessment measures used for evaluating the classification performance. We then brief out the effect of the parameters on the classification algorithm and finally compare the classification performance of hyperspectral image before and after applying spatial preprocessing step.

6.1. Dataset Description

The hyperspectral image used in our experiments is the Indian Pines image shown in Fig. 3(a), acquired by Airborne Visible/IR Imaging Spectrometer (AVIRIS) sensor. This image contains 145×145 pixels and 220 spectral bands in the spectral range of 400-2500nm. Noisy bands 104-108, 150-163, 220 were removed due to water absorption. Hence, the processed image has 145×145 pixels and 200 spectral bands. Fig. 3(b) shows the ground truth available for the Indian Pines image consisting of 16 classes, of which ten correspond to different kind of crops, five correspond to vegetation, and one corresponds to a building.



(a)



(b)

Figure 3. AVIRIS data scene a) color composite image, b) ground data for the scene and description of the classes

6.2. Accuracy Assessment

Classification accuracy in case of hyperspectral images is best expressed by using confusion matrix or error matrix. The major diagonal of the error matrix represents the properly classified land use categories. The non-diagonal elements of the matrix represent errors of omission or commission. Omission errors correspond to non diagonal row elements and commission errors are represented by non diagonal column elements. The classwise accuracy, average accuracy, and overall accuracy is measured from the confusion matrix as given below

$$Classwise Accuracy = \frac{\text{Correctly classified pixels in each class}}{\text{Total number of pixels in each class}}$$

$$Average Accuracy = \frac{\text{Sum of the accuracies of each class}}{\text{Total number of class}}$$

$$Overall Accuracy = \frac{\text{Total number of correctly classified pixels}}{\text{Total number of pixels}}$$

6.3. Results and Discussion

In this section, the Perona-Malik diffusion algorithm is applied to a real hyperspectral image, and is then evaluated quantitatively and visually using classification maps. Briefly explaining our work, we first perform Perona-Malik non linear diffusion on each band of Indian Pines image after which the hyperspectral cube is classified using joint sparsity model. The performance of the proposed method is then compared with the sparsity based classification of hyperspectral image without applying spatial preprocessing method.

The effect of spatial preprocessing on the hyperspectral image classification depends on the parameters such as number of iteration, step size and diffusion control parameter k value. We find that as the number of iteration increases, the entire image gets blurred. This is one of the drawbacks of diffusion techniques. Over-diffusion of the image may result the new smoothed image to be too far from the original version. Hence, the number of iterations is to be limited so that the features are preserved. Here, the number of iteration is limited to 3, where we find the noise been attenuated while the features been preserved. The step size plays an important role in the stability of the diffusion process. The step size should be chosen less than 0.25 in order to result in a stable solution scheme. Diffusion control parameter k influences the smoothing process and mainly depends on the hyperspectral image used. Here, we have set the k value as .012.

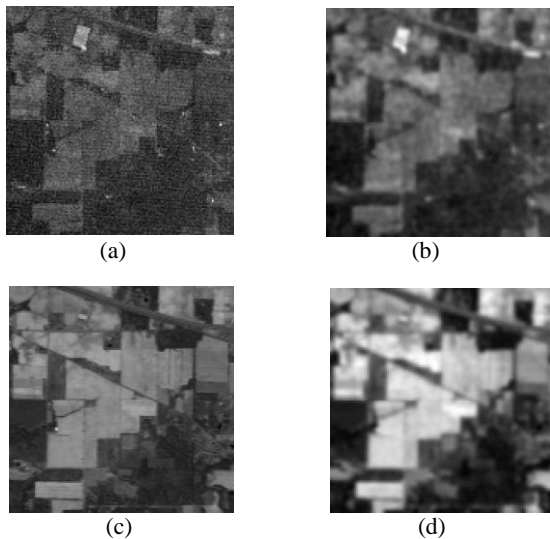


Figure 4. Effect of preprocessing (a) original band 2 (b) band 2 after preprocessing (c) original band 145 (d) band 145 after preprocessing

Fig. 4 shows the effect of Perona-Malik diffusion. Comparing the original and preprocessed version, we find that in noisy band 2, noise has been reduced with most of the features been preserved whereas the noiseless band 145 has been smoothed out.

In order to observe the effect of Perona-Malik diffusion on the classification of the real hyperspectral image, we present the classification maps obtained using Indian Pines image before preprocessing and after preprocessing, which are depicted in Fig. 5. According to the visual analysis, the proposed method gives more accurate classification map than the classification map obtained without applying the spatial preprocessing technique. Furthermore from Table 3-4, it is apparent that preprocessed image produces higher classification accuracy.

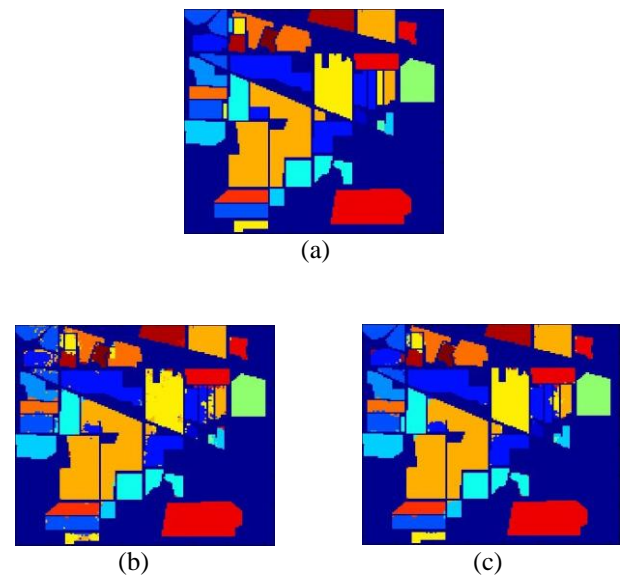


Figure 5. (a) Ground truth map. Classification maps obtained using Indian Pines image (b) before preprocessing (c) after preprocessing (proposed method)

Finally, a main issue to be highlighted from the results is that an optimal sparsity level K and window size acts as important parameters for varying the classification accuracy. In this work, we have set the window size as 9×9 and sparsity level K as 30, which gives highest classification accuracy as shown in Table 1. The sparsity level K is varied from 10 to 30 and the window size is varied from 3×3 to 9×9 . We find that as the window size increases, accuracy increases. This

Table 1. Effect of window size and Sparsity level K on Overall Accuracy of Indian Pines image before preprocessing

Window size	Sparsity Level K	Overall Accuracy
3×3	10	92.34
	20	83.66
	30	82.71
5×5	10	90.47
	20	90.87
	30	90.93
7×7	10	89.30
	20	94.02
	30	94.03
9×9	10	88.57
	20	94.24
	30	94.77

is because of the fact that the Indian Pines image has large homogenous regions. We also find that when the window size is large, a small sparsity level gives reduced accuracy results since the pixels in the neighbourhood cannot be faithfully approximated using a few training samples. Also for a large sparsity level, the classification accuracy increase as the window size increases [3].

7. Conclusion

In this paper, we discussed the overall impact of spatial preprocessing technique namely Perona-Malik, a non-linear diffusion scheme applied prior to sparsity based hyperspectral image classification. We considered the hyperspectral imagery as two-dimensional images and applied the diffusion scheme on each band of the hyperspectral data. The proposed method was tested on a real hyperspectral dataset. The experimental results proved that the spatial preprocessing technique applied on the hyperspectral band improved the quality of image and provided better within-class variations which lead to the increase in classification accuracy.

8. References

- [1] J. Wright, A. Y. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210-227, Feb. 2009.
- [2] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Sparse representation for target detection in hyperspectral imagery," *IEEE Journal of Selected Topics in Signal Processing* vol. 5, no. 3, pp. 629-640, June 2011.
- [3] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," In *IEEE Trans. On Geoscience and Remote Sensing*, no. 10, pp. 3973-3985, Oct. 2011.
- [4] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Simultaneous joint sparsity model for target detection in hyperspectral imagery," In *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 4, pp. 676-680, July 2011.
- [5] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," In *Signal Processing, special issue on Sparse approximations in signal and image processing*, vol. 86, pp. 572-588, Mar. 2006.
- [6] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," In *IEEE Trans. on Geoscience and Remote Sensing*, vol. 43, no. 6, pp. 1351-1362, June 2005.
- [7] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral Image Classification via Kernel Sparse Representation," In *IEEE Trans. on Geoscience and Remote Sensing*, 2011.
- [8] U. Srinivas, Y. Chen, V. Monga, N. M. Nasrabadi, and T. D. Tran, "Exploiting sparsity in hyperspectral image classification via graphical models," In *IEEE Geoscience and Remote Sensing Letters*, 2011.
- [9] V. M. Patel and R. Chellappa, "Sparse Representations, Compressive Sensing and Dictionaries for Pattern Recognition," In *Asian Conference on Pattern Recognition (ACPR)*, Beijing, China, 2011.
- [10] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629-639, Jul. 1990.
- [11] S. K. Weeratunga and C. Kamath, "A Comparison of PDE-based Nonlinear Anisotropic Diffusion Techniques for Image Denoising," *Image Processing: Algorithms and Systems II, SPIE Electronic Imaging, Santa Clara*, January 2003.
- [12] David Hwang, "Anisotropic Diffusion of Hyperspectral Imagery," June 2009.
- [13] Yi Wang, Ruiqing Niu, and Xin Yu, "Anisotropic diffusion for hyperspectral imagery enhancement," *IEEE Sensors Journal*, vol. 10, no. 3, March 2010.
- [14] Santiago Velasco-Forero and Vidya Manian, "Improving Hyperspectral Image Classification Using Spatial Preprocessing," *IEEE Geoscience And Remote Sensing Letters*, Vol. 6, No. 2, April 2009.
- [15] Ping Liang and Y. F. Wang, "Local Scale Controlled Anisotropic Diffusion with Local Noise Estimate for Image Smoothing and Edge Detection".
- [16] AVIRIS NW Indiana's Indian Pines 1992 Data Set. [Online]. Available: <http://cobweb.ecn.purdue.edu/biehl/MultiSpec/documentation.html>

Table 2. Confusion Matrix for Indian Pines image calculated based on our proposed method

		CLASSIFICATION RESULTS																
GROUND TRUTH		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
	1	50	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	1379	2	0	0	2	0	0	0	10	26	0	0	1	14	0	
	3	0	5	808	5	5	1	0	0	0	0	0	6	4	0	0	0	
	4	0	0	0	232	0	0	0	0	0	0	0	2	0	0	0	0	
	5	0	0	7	0	475	0	0	0	0	10	5	0	0	0	0	0	
	6	0	0	0	0	0	743	0	0	0	0	2	0	0	2	0	0	
	7	0	0	0	0	18	0	8	0	0	0	0	0	0	0	0	0	
	8	0	0	0	0	0	0	0	489	0	0	0	0	0	0	0	0	
	9	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	
	10	0	18	4	0	0	0	1	0	0	926	8	3	0	8	0	0	
	11	0	6	0	0	0	10	0	1	0	5	2441	0	1	2	2	0	
	12	0	0	7	2	0	0	0	0	0	1	0	599	0	0	1	4	
	13	0	0	0	0	2	0	0	0	0	0	0	0	210	0	0	0	
	14	0	0	0	0	0	0	0	1	0	0	4	0	0	1284	5	0	
	15	0	0	0	0	0	0	0	0	0	5	0	0	0	0	375	0	
	16	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	91	

Table 3. Comparison of classwise accuracy of Indian Pines image before and after applying spatial preprocessing (proposed method)

Class	before applying spatial preprocessing	after applying Spatial preprocessing (proposed method)
1	85.18519	92.59259
2	91.98047	96.16457
3	92.20624	96.88249
4	94.01709	99.1453
5	94.3662	95.57344
6	98.92905	99.46452
7	50	30.76923
8	100	100
9	5	0
10	86.77686	95.66116
11	98.29822	98.906
12	88.59935	97.557
13	98.58491	99.0566
14	99.53632	99.2272
15	95	98.68421
16	95.78947	95.78947
Overall	94.77	97.53
Average	85.89	87.217