

ADMM based Hyperspectral Image Classification Improved by Denoising using Legendre Fenchel Transformation

C. Aswathy*, V. Sowmya and K. P. Soman

Centre for Excellence in Computational Engineering and Networking, Amrita Vishwa Vidyapeetham, Coimbatore - 641 112, Tamil Nadu, India; aswathy0257@gmail.com, vsowmya@cb.amrita.edu

Abstract

This paper discusses about a sparsity based algorithm used for Hyperspectral Image (HSI) classification where the test pixel vectors are sparsely represented as the linear combination of a few number of training samples from a well-organised dictionary matrix. The sparse vector is obtained using Basis Pursuit (BP) which is a constrained ℓ_1 minimization problem. This problem is solved by using a simple and powerful iterative algorithm known as Alternating Direction Method of Multipliers (ADMM) which significantly reduces the computational complexity of the problem and thereby speeds up the convergence. The classification accuracy is considerably improved by including efficient preprocessing techniques to remove the unwanted information (noise) present in Hyperspectral images. This paper uses a fast and reliable denoising technique based on Legendre Fenchel Transformation (LFT) to effectively denoise each band of HSI prior to ADMM based classification (proposed method). A comparison of proposed technique with one of the convex optimization tools namely, CVX is given to exhibit the fast convergence of the former method. The experiment is performed on standard Indian Pines dataset captured using AVIRIS sensor. The potential of the proposed method is illustrated by analyzing the classification indices obtained with and without applying any preprocessing methods. With only 10% training set, an overall accuracy of 96.76% is obtained for the proposed method at a much faster rate compared to computation time taken by CVX solver.

Keywords: Alternating Direction Method of Multipliers, Basis Pursuit, Classification, Hyperspectral Denoising, Legendre Fenchel Transformation

1. Introduction

The generous information available from the highly correlated spectral bands of Hyperspectral Images (HSI) are used for countless applications in the areas of agriculture, defence, medical science, urban planning, forestry and monitoring. The availability of these abundant spectral and spatial information and advancement in the field of computational engineering has paved the way for the development of several specialised algorithms to exploit the non linear capability of HSI. These algorithms are widely used in hyperspectral data analysis methods namely, image classification, unmixing, pansharpening, target detection etc.

Hyperspectral image classification is one of the major areas of research in the field of remote sensing where a number of new classifiers are introduced in recent years. Support Vector Machines (SVM)¹, Minimum Spanning Forest (MSF)², Probability based Multinomial Logistic Regression (MLR)³ etc. are some of the commonly used classifiers for HSI classification which show good classification results. Despite these, several sparsity based algorithms are also utilised for the classification of HSI. Yi Chen et al.⁴ have introduced a novel sparsity based algorithm which shows that the pixels in HSI can be sparsely represented as the linear combination of a very few number of training samples. Generally, an optimization based classification is addressed by

* Author for correspondence

various solvers. CVX⁵ is one of such modeling system available for solving the convex optimization problems. The problems involving massive optimization and high dataset dimension utilizes an efficient algorithm namely, Alternating Direction Method of Multipliers (ADMM) which can compete with other well known methods⁶.

Since the hyperspectral images are highly prone to noise, the ADMM based classification is aided by certain preprocessing techniques which improve the overall accuracy obtained during HSI classification. An adaptive total variation denoising method was proposed by Yuan et al.⁷ for the noise reduction process utilising both spatial and spectral noise differences. Guangyi Chen et al.⁸ proposed a method using Principle Component Analysis (PCA) to decorrelate the required information of the hyperspectral data cube from the noise present in the images. They also used 1-D dualtree complex wavelet transform and 2-D bivariate wavelet thresholding to remove noise in the spectrum of each pixel and low energy PCA output channels respectively. Denoising based on Perona Malik diffusion⁹, wavelet decompositions and sparse approximation methods¹⁰ are some other existing preprocessing techniques for effective HSI denoising. The Gaussian noise present in the colour remote sensing image is effectively removed by using a denoising technique based on Legendre Fenchel Transformation¹¹.

In this paper, deploying a fast and reliable Legendre Fenchel ROF denoising as spatial preprocessing step prior to ADMM based classification has made a significant improvement in the classification indices of HSI classification. Legendre Fenchel Transformation utilizes the concept of duality where the function in the primal space is mapped to the dual space for the better understanding of the problem. The use of simple and powerful Alternating Direction Method of Multipliers (ADMM) algorithm instead of CVX solver in HSI classification problem helps to significantly reduce the computation time needed for the experiment. Rest of the sections in this paper is arranged as follows. Denoising technique based on Legendre Fenchel Transformation is described in section 2. ADMM based HSI classification is given in section 3. Section 4 presents the proposed method. Section 5 discusses about experimental results and analysis. The paper is concluded in section 6

2. HSI Denoising using LFT

The 3-dimensional hyperspectral data cube is of the form $X \in \mathcal{R}^{n_1 \times n_2 \times n_3}$ where n_b and $n_1 \times n_2$ represent the number of bands and the number of pixels respectively. Various preprocessing algorithms are used to remove the noise present in different bands of the HSI. Such a denoised image should preserve the sharp edges of the original image while reducing the noise content and must have the same statistical properties as that of the original noisy image. These objectives are acquired by using an effective denoising method based on Legendre Fenchel Transformation.

The concept of duality is utilised in LFT where the primal space $(x, f(x))$ is mapped to dual space $(p, f^*(p))$ to better represent the problem¹². For a continuous function, $f: \mathcal{R} \rightarrow \mathcal{R}$, LFT is given as

$$f^*(p) = \sup_{x \in \mathcal{R}} \{px - f(x)\} \quad (1)$$

where x and p represents the primal variable and dual variable respectively.

2.1 LFT for Dual Formulation of ROF Model

Unlike the Euler Lagrange ROF model¹³, the ROF formulation using LFT consumes very less computation time which makes it one of the best denoising techniques existing in recent days. The use of nabla matrix in the calculation of gradient and divergence has considerably reduced the computational complexity of the problem. Therefore, denoising using Legendre Fenchel Transformation (LF-ROF denoising) is a fast and reliable method compared to other existing methods¹¹. The standard form of dual ROF model is given as:

$$\min_{u \in X} \|\nabla u\|_1 + \frac{\lambda}{2} \|u - g\|_2^2 \quad (2)$$

By using Legendre Fenchel Transformation,

$$\|\nabla u\|_1 = \max_{p \in P} \left(\langle p, \nabla u \rangle - \delta_p(P) \right) \quad (3)$$

where P is the indicator function. Dual formulation of ROF by using LFT is given as,

$$\min_{u \in X} \max_{p: \|p\|_\infty \leq 1} \left(\langle p, \nabla u \rangle - \delta_p(P) + \frac{\lambda}{2} \|u - g\|_2^2 \right) \quad (4)$$

The primal and dual updates are given¹¹ as,

$$u^{n+1} = \frac{u^n + \tau \operatorname{div} p^{n+1} + \tau \lambda g}{1 + \tau \lambda} \quad (5)$$

$$p^{n+1} = \frac{p^n + \sigma \nabla u^n}{\max(1, |p^n + \sigma \nabla u^n|)} \quad (6)$$

3. Hyperspectral Image Classification

The supervised hyperspectral image classification involves both training phase and testing phase. Randomly selected data feature vectors are given in the training phase and the rest of the sample or the image as a whole is given for the testing phase. Class labels of the training samples are known, whereas the class labels of the test pixel are verified using the ground truth information.

In the sparsity based classification problem, a hyperspectral data of B dimension with N feature (spectral) vectors is considered. The dictionary matrix A of size $B \times D$ where $D \subset N$ is formed using the training set by concatenating the randomly selected pixels from each class. The dictionary is represented as $A = [A_1, \dots, A_M]$, where the subdirectory is given by $A_i = [a_{i1} \ a_{i2} \ \dots \ a_{im_i}]$. A_i represents the set of training data vectors belonging to i^{th} class where $i \in M$, M is the total number of classes and m_i is the total number of training pixels belonging to a particular class.

A test pixel of a particular class can be represented as the linear combination of training pixels of the same class which is represented as⁴,

$$y = \sum_{k=1}^K A_k x_k \quad (7)$$

where K is the number of non zero elements considered in the sparse vector x (sparsity level). An approximation of test pixel vector belonging to class i is given by,

$$\hat{y} = \alpha_1 a_1 + \alpha_2 a_2 + \dots + \alpha_{m_i} a_{m_i} \quad (8)$$

where α_i are the weights for a_i belonging to A_i , $i \in M$ values are the non zero elements in the sparse vector x . The labelling of test vector is done using the residue vectors obtained from the sparse estimate \hat{y} for each class. The test pixel vector is assigned to the class with minimum residue. The coefficient for the above mentioned linear

combination is given by the sparse vector x from the system of equation $Ax = y$. This optimization problem is solved using the equation,

$$\hat{x} = \arg \min \|x\|_0 \quad (9)$$

$$Ax = y$$

The above NP-hard problem is replaced by a linear problem which uses Basis Pursuit, a constrained l_1 minimization problem used to obtain sparsity based classification⁴. This experiment uses a fast and flexible algorithm called Alternating Direction Methods of Multipliers (ADMM) to solve the BP formulation which takes lesser computation time compared to convex optimization solver like CVX.

3.1 ADMM based Classification

The sparsity based classification involves two major steps. One is to obtain the sparse vector x using Basis Pursuit addressed by ADMM and the other is residue calculation to assign the class label of the test pixel. Basis Pursuit (BP) uses l_1 minimization to find out the sparse linear combination of spectra from large libraries¹⁴. The solution for the BP is obtained via ADMM which is a simple and powerful iterative algorithm for distributed convex optimization problems. ADMM comprises the benefits of both augmented Lagrangian techniques and dual decomposition⁶. In ADMM, a difficult problem is decomposed into sequence of simpler problems. The presence of Lagrangian multiplier in ADMM algorithm helps to reach the solution faster since it speeds up the convergence process¹⁵.

ADMM based classification algorithm is given as:

- The dictionary matrix $A = [A_1, \dots, A_M]$ of size $B \times D$ is formed by randomly selecting the training samples from each class. Each subdirectory of dictionary A is given by $A_i = [a_{i1} \ a_{i2} \ \dots \ a_{im_i}]$.
- The columns of dictionary matrix A are normalized.
- Using BP solved via ADMM, the sparse vector x is determined.
- Objective function of Basis Pursuit is given as⁶,

$$\hat{x} = \arg \min \|x\|_1 \quad (10)$$

$$Ax = y$$

with variable $x \in R^n$, $A \in R^{m \times n}$ and $y \in R^m$ with $m < n$. l_0 optimization is used for obtaining the sparse solution is non convex and hence it is difficult to solve.

- Above equation in ADMM format is rewritten as,

$$\min f(x) + \|z\|_1 \tag{11}$$

subject to $x - z = 0$

where $f(x)$ is an indicator function of $\{x \in R^n \mid Ax = b\}$

- The update corresponding to x^{k+1} , z^{k+1} and u^{k+1} are obtained using ADMM.

Augmented Lagrangian corresponds to the minimization problem is given by,

$$L_\rho(x, z, u) = f(x) + \|z\|_1 + \frac{\rho}{2} \|x - z + u\|_2^2 \tag{12}$$

x update involves solving a linearly constrained minimum Euclidean norm problem.

- The concept of projection is used to find the update for x . i.e.,

where Π is the projection onto $\{x \in R^n \mid Ax = b\}$ i.e., projecting $x^k = (z^k - u^k)$ onto $Ax = y$.

- The update for x is given as⁶,

$$x^{k+1} = (I - A^T(AA^T)^{-1}A)(z^k - u^k) + A^T(AA^T)^{-1}y \tag{13}$$

- The update for z and u is represented as,

$$z^{k+1} = S_{1/\rho}(x^{k+1} + u^k) \tag{14}$$

where $S_{1/\rho}$ is the soft thres holding operator.

- The new dictionary matrix A_{new} is obtained by iterating through the number of classes M such that the columns other than those belonging to iteration value (class) are set to zero.

- The residue r_i is found as,

$$r_i = \|y - A_{new}^i * x\|_2 \tag{15}$$

- The residue for test vector is determined over all classes.
- The test pixel vector is assigned to the class with minimum residue. This is mathematically expressed as,

$$Class(y) = \arg \min_{i=1, \dots, M}(r_i) \tag{16}$$

This algorithm is fast and provides satisfactory classification output.

4. Proposed Method

The proposed method involves the sparsity based HSI classification with constrained l_1 optimization addressed by ADMM, which is aided by an effective preprocessing step called Legendre Fenchel ROF denoising. The experiment is performed on standard AVIRIS Indian Pines dataset. Spatial preprocessing and classification are the two major steps in the proposed method. Figure 1 represents the flow graph of the proposed method.

4.1 Spatial Preprocessing

HSI are highly prone to noise that degrades the quality of the captured image and thereby reduce the values of classification indices like classwise accuracy, overall accuracy and average accuracy. An efficient spatial preprocessing is performed to overcome this challenge. This experiment uses denoising based on Legendre Fenchel Transformation (LF-ROF denoising) to effectively denoise each band of HSI without losing the edge information.

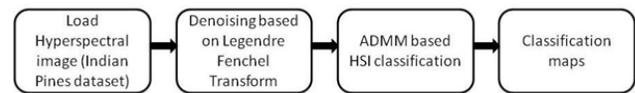


Figure 1. Proposed method.

4.2 Classification using Basis Pursuit

During classification, the pixels present in the HSI are separated to training and testing samples. Generally, during the training phase, 10%, 20%, 30% or 40% samples are randomly selected from each class to form the dictionary matrix and rest of the samples (or whole samples) are given for testing. Classification accuracy increases with increase in the percentage of training samples. In this experiment, classification involves finding the sparse vector x using Basis Pursuit which is solved by ADMM solver and assigning the class labels to the test pixel vectors using the residue.

5. Experimental Result Analysis

This section gives the description about the dataset used for the present experiment, accuracy assessment measures, analysis of the effect of preprocessing on classification of HSI and comparison of CVX based and ADMM based classification.

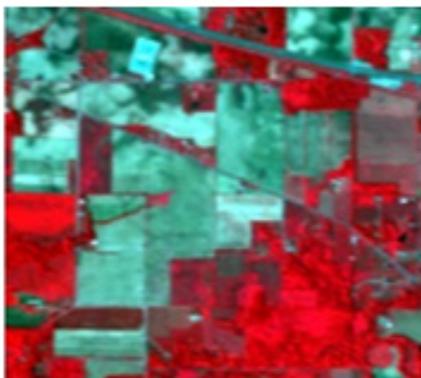
5.1 Dataset Description

Indian Pines is one of the standard datasets available which is captured using AVIRIS sensor system on June 12, 1992 over a 2 x 2 mile portion of Northwest Tippecanoe, Indiana is used for the experimental study. It possesses 224 contiguous spectral bands over a range of 400-2500 nano metres. Four out of 224 bands are removed since it does not have any useful information. Further the number of bands is reduced to 200 by eliminating the water absorption bands from 104-108, 150-163 and 220. Each band possesses 145x145 pixels. The dataset is with a spatial resolution of 20 m per pixel, spectral resolution of 10 nm and radiometric resolution of 16 bit. Ground truth data of Indian Pines dataset consists of 16 classes. Figure 2(a) and Figure 2(b) show the colour image and ground truth image together with class descriptions of AVIRIS Indian Pines data scene.

5.2 Accuracy Assessment Measures

As a part of accuracy assessment measures, both subjective and objective assessments are performed. Subjective assessment is done through visual interpretation while objective measure involves the comparison of the classification maps generated and the ground truth (reference) data. Another method of accuracy assessment is the computation of confusion matrix which is used to find the overall, classwise and average accuracies⁹. Another index used to quantify the agreement of classification is Kappa coefficient.

$$OA = \frac{\text{Total number of correctly classified pixels}}{\text{Total number of pixels}} \quad (17)$$



(a)

$$AA = \frac{\text{Sum of accuracies of each class}}{\text{Total number of class}} \quad (18)$$

$$CA = \frac{\text{Correctly classified pixels in each class}}{\text{Total number of pixels in each class}} \quad (19)$$

$$\text{Kappa Coefficient} = (N * A - B) / (N^2 - B) \quad (20)$$

where N is the total number of pixels, A is the number of correctly classified pixels, B is the sum of product of row and column in confusion matrix.

OA : Overall Accuracy, AA : Average Accuracy, CA : Classwise Accuracy

6. Results and Discussions

The experiment is conducted on standard AVIRIS Indian Pines dataset to evaluate the performance of optimization based HSI classification. To improve the classification indices of such classifications, an effective denoising technique using Legendre Fenchel Transformation is also included prior to classification of HSI. The role of other existing preprocessing techniques like Total Variation (Euler Lagrange ROF) denoising and wavelet based denoising are also considered to exhibit the importance of LF-ROF denoising. The effect of various denoising techniques on HSI are analysed by taking band 165 of the Indian Pines dataset as a sample band. Figure 3 (a)–(d) shows the Noisy band 165, preprocessing on band 165 using TV denoising, Wavelet based denoising and LF-ROF denoising. Both TV denoising and wavelet based denoising fail to preserve sharp edges in the image whereas LF-ROF denoising helps in smoothing the noisy image while preserving the edge information. LF-ROF



(b)

Figure 2. AVIRIS Indian Pines dataset. (a) Colour composite image. (b) Ground truth image.

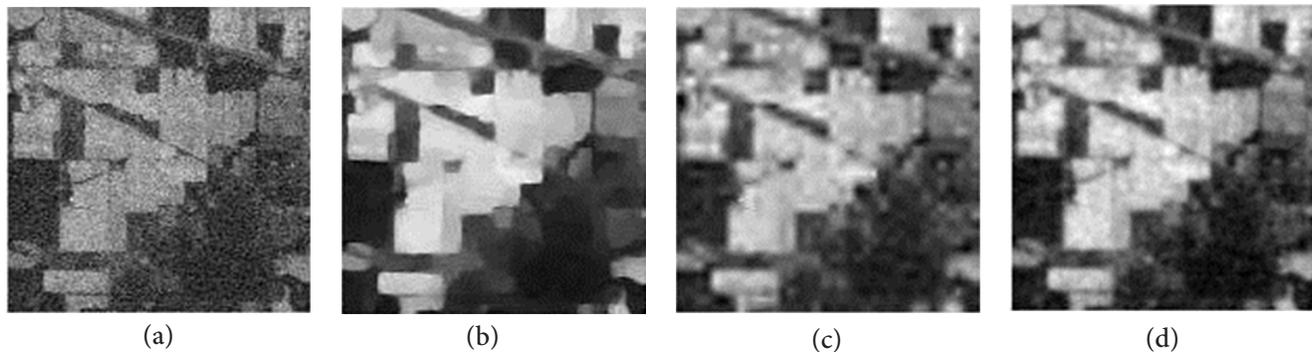


Figure 3. Preprocessing on Indian Pines dataset (Band 165) using different denoising methods. (a) Original image. (b) TV Denoising. (c) Wavelet based Denoising. (d) LF-ROF denoising.

denoising is a fast and reliable method of denoising whose performance relies on the value of control parameter, number of iteration and Lipchitz constant selected experimentally. Each denoised band of the hyperspectral dataset is stacked to form a 3-D hyperspectral data cube which is given for optimization based classification.

This optimization based classification process involves a constrained l_1 minimization problem (BP) where the sparse vector x is obtained by various optimization solvers like ADMM and CVX. The performances of the proposed method involving ADMM aided by LF-ROF denoising is evaluated based on time and computational complexity. The evaluation shows the effectiveness of ADMM over the convex optimization solver, CVX. Table II provides the time comparison between the two solvers (ADMM and CVX) used in HSI classification.

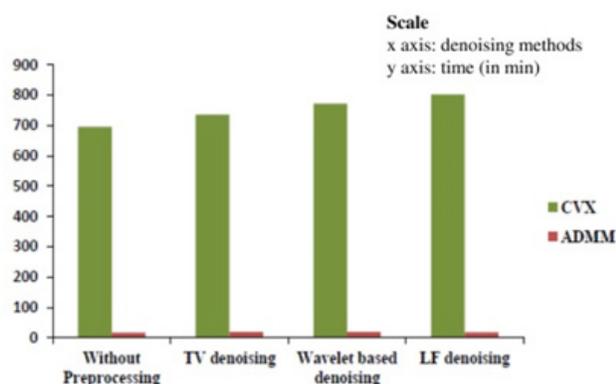


Figure 4. Graphical representation of Time Analysis of HSI Classification.

The experiment is conducted on an Intel(R) Core(TM) i7-4790S CPU @ 3.2 GHz, 64 bit operating system with 8 GB RAM. High processing time is required for the complex operations performed on large dimensional data

like HSI. By the use of powerful algorithms like ADMM, the computational time is greatly reduced. On analysing the computational time given in Table 2, it is clear that ADMM based classification is much faster compared to CVX solver in complex computational problems. Hence it is inferred that CVX tool is not suitable for very large problems. On the other hand, ADMM is a simple and powerful algorithm which is faster and flexible compared to other state-of-art methods. Graphical representation of Table 2 is given in Figure 4.

Table 2. Time Analysis of HSI classification (10%training samples)

Sl.No.	Various preprocessing techniques	Time taken for classification of Indian Pines Dataset (in min)	
		CVX	ADMM
1	Without Preprocessing	695.69	17.81
2	TV (Euler Lagrange ROF) denoising	736.26	18.89
3	Wavelet based denoising	772.04	19.00
4	Legendre Fenchel ROF denoising	882.50	18.45

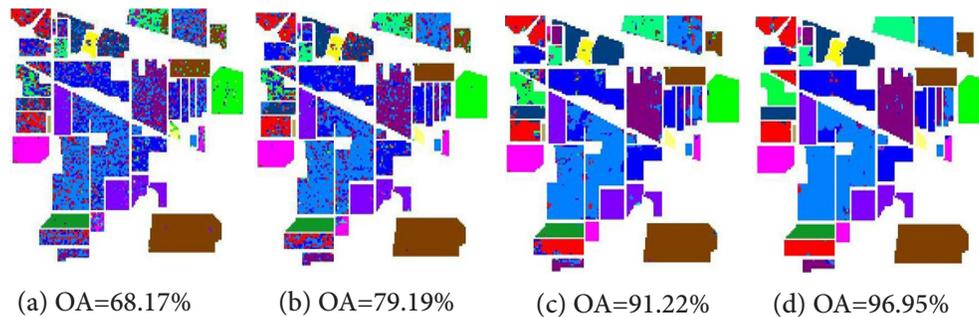
During classification, training and testing samples for the dataset is generated from the ground truth information. The training set is generated by the random selection of pixels from each class and the remaining pixels are used for testing purpose. Table 1 shows the effect of various preprocessing techniques on CVX and ADMM based hyperspectral image classification. 10% samples are selected randomly from each class of Indian Pines Dataset for training purpose. Analysis of Table 1 shows the importance of preprocessing of HSI prior to classification, by comparing the classification accuracies obtained before and after preprocessing. It is inferred that

Table 1. Comparison of CVX and ADMM based classification results obtained without and with preprocessing on Indian Pines dataset

Class	Class Name	Without		TV denoising		Wavelet based		LF denoising	
		Preprocessing		CVX	ADMM	CVX	ADMM	CVX	ADMM
1	Alfalfa	69.57	80.43	97.83	97.83	100.00	97.83	100.00	100.00
2	Corn-notill	57.14	57.35	78.01	74.02	91.67	90.20	95.31	94.75
3	Corn-mintill	49.76	42.29	60.72	52.65	79.16	84.22	95.18	90.36
4	Corn	32.07	41.35	44.30	60.76	88.19	84.39	100.00	97.47
5	Grass-pasture	78.88	76.60	90.27	90.48	96.89	96.69	97.72	98.14
6	Grass-trees	92.05	87.53	94.11	94.25	98.08	97.95	99.59	99.73
7	Grass-pasture	92.86	82.14	100.00	100.00	100.00	100.00	100.00	100.00
8	mowed	94.98	91.63	99.79	98.54	99.37	99.79	100.00	100.00
9	Hay-windrowed	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
10	Oats	56.07	48.97	72.74	58.23	87.76	88.07	94.44	93.72
11	Soybean-notill	61.22	61.30	72.26	79.27	86.19	89.12	95.56	97.23
12	Soybean-mintill	58.68	54.13	70.32	71.16	95.95	94.27	98.99	96.46
13	Soybean-clean	99.02	98.05	99.51	98.54	97.56	98.54	100.00	100.00
14	Wheat	95.10	95.49	98.02	96.76	98.89	96.60	99.60	99.92
15	Woods	54.40	57.51	70.73	73.58	88.86	91.97	95.85	97.67
16	Buildings-Grass-Trees-Drives Stone-Steel-Towers	91.40	94.62	91.40	90.32	95.70	95.70	97.85	100.00
	Overall Accuracy	68.17	66.50	79.19	78.59	91.22	91.81	96.95	96.76
	Average Accuracy	73.95	73.09	83.75	83.52	94.02	94.08	98.13	97.84
	Kappa	0.6369	0.6172	0.7626	0.7247	0.9001	0.9068	0.9652	0.9630

the proposed method using ADMM based classification aided by LF-ROF denoising outperforms classification with other existing preprocessing methods. Since both CVX and ADMM are used to solve the same convex optimization problem (BP), they provide almost similar accuracies results for the classification problem. CVX based classification gives an overall classification accuracy of 68.17% without applying any preprocessing, 79.19% with TV denoising, 91.22% with wavelet denoising and 96.95% with the LF-ROF denoising which are shown in Figure 5(a)–(d) respectively. Whole samples or the

samples other than the training samples are given for the testing purpose. The test set (Ground truth) is given in Figure 6(a). Classification accuracies obtained by ADMM are given in Figure 6(b)–(d), i.e., 66.50% without preprocessing, 78.79% with TV denoising and 91.81% with wavelet based denoising. By taking only 10% of training samples, the proposed method attains an overall classification accuracy of 96.76% in very less computation time compared to CVX based classification. The results obtained are far better compared to the overall accuracies obtained by using other existing preprocessing techniques.

**Figure 5.** Classification maps of CVX based classification. (a) Without Preprocessing, (b) With TV denoising, (c) With Wavelet Denoising, (d) With Legendre Fenchel Denoising.

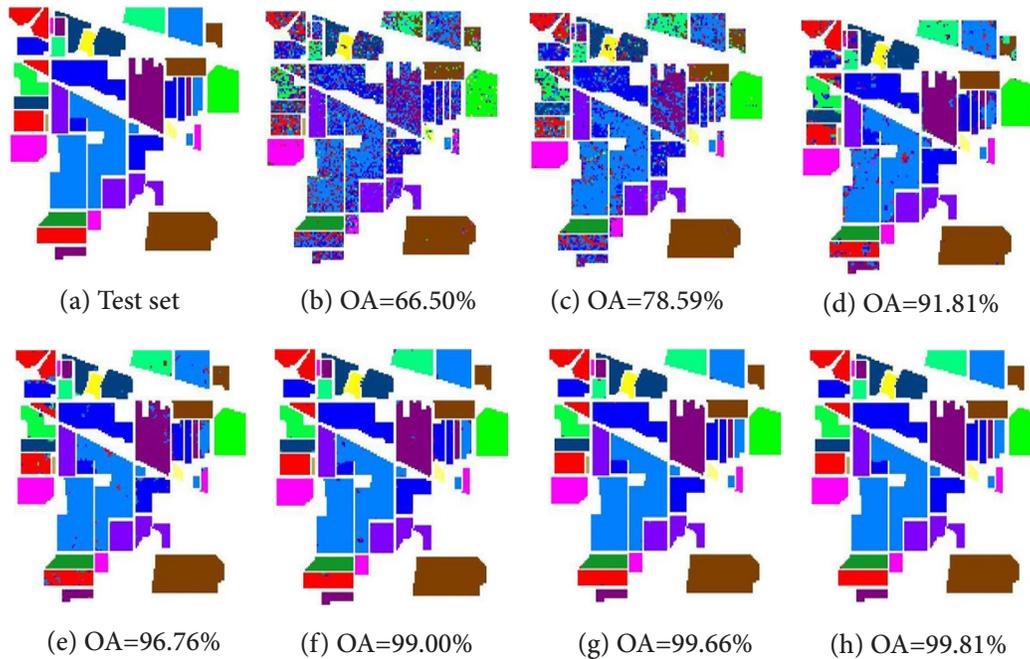


Figure 6. (a) Test Set (Ground truth). Classification maps of ADMM based classification: (b) Without Preprocessing, (c) With TV denoising, (d) With Wavelet Denoising, (e) With Legendre Fenchel Denoising(10% training samples), (f) With Legendre Fenchel Denoising(20% training samples), (g) With Legendre Fenchel Denoising(30% training samples), (h) With Legendre Fenchel Denoising(40% training samples).

Classification accuracy can be improved by increasing the number of samples given for training. Improvement in the classification indices of proposed method along with increase in the training set (10%, 20%, 30% and 40%) is depicted in Table 3. With 40% training set, the proposed method provides an overall classification accuracy of 99.81%. The performance of proposed method with different training sets (10%, 20%, 30% and 40%) is given in Figure 6 (e)–(h) respectively.

7. Conclusion

This paper discusses the role of ADMM in solving sparsity based constrained l_1 optimization problem (BP) for classification of Hyperspectral Images. The computation efficiency of ADMM is proved by comparing its time elapsed for HSI classification with that of CVX optimization tool. The classification indices are further improved by a fast and efficient preprocessing technique using Legendre Fenchel Transformation that effectively denoised each band of the HSI preserving the

edge information. Experimental analysis proves that the proposed method outperforms other existing techniques.

Table 3. Classification results of ADMM based classification for different training sets

Percentage of training for ADMM classification with LF denoising	10	20	30	40
Overall Accuracy	96.76	99.00	99.66	99.81
Average Accuracy	97.84	99.26	99.68	99.78
Kappa Coefficient	0.9630	0.9887	0.9961	0.9979

8. Acknowledgement

The authors would like to thank Mr. Nidhin Prabhakar, Research Associate, Centre for Excellence in Computational Engineering and Networking, Amrita Vishwa Vidyapeetham, for his valuable support. The authors also wish to convey sincere gratitude to all who directly or indirectly involved in the progress of this manuscript.

9. References

1. Gualtieri JA, Chettri SR, Cromp RF, Johnson LF. Support vector machine classifiers as applied to AVIRIS data. Proc Eighth JPL Airborne Geoscience Workshop. Citeseer; 1999 Feb.
2. Bernard K, Tarabalka Y, Angulo J, Chanussot J, Benediktsson JA. Spectralspatial classification of hyperspectral data based on a stochastic minimum spanning forest approach. *IEEE Transactions on Image Processing*. 2012 Apr; 21(4):2008–21.
3. Li J, Bioucas-Dias JM, Plaza A. Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning. *IEEE Transactions on Geoscience and Remote Sensing*. 2010 Nov; 48(11):4085–98.
4. Chen Y, Nasrabadi NM, Tran TD. Hyperspectral image classification using dictionary-based sparse representation. *IEEE Transactions on Geoscience and Remote Sensing*. 2011 Oct; 49(10):3973–85.
5. Grant M, Boyd S, Ye Y. CVX: Matlab software for disciplined convex programming. 2015 Jun.
6. Boyd S, Parikh N, Chu E, Peleato B, Eckstein J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*. 2011; 3(1):1–122.
7. Yuan Q, Zhang L, Shen H. Hyperspectral image denoising employing a spectralspatial adaptive total variation model. *IEEE Transactions on Geoscience and Remote Sensing*. 2012 Oct; 50(10):3660–77.
8. Chen G, Qian SE. Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage. *IEEE Transactions on Geoscience and Remote Sensing*. 2011 Mar; 49(3):973–80.
9. Balakrishnan K, Sowmya V, Soman KP. Spatial preprocessing for improved sparsity based hyperspectral image classification. *International Journal of Engineering Research and Technology*. ESRSA Publications 2012 Jul; 1(5).
10. Zelinski AC, Goyal VK. Denoising hyperspectral imagery and recovering junk bands using wavelets and sparse approximation. *IEEE International Conference on Geoscience and Remote Sensing Symposium (IGARSS 2006)*; Denver, CO. 2006 Jul 31-Aug 4. p. 387–90.
11. Santhosh S, Abinaya N, Rashmi G, Sowmya V, Soman KP. A novel approach for denoising coloured remote sensing image using Legendre Fenchel Transformation. *International Conference on Recent Trends in Information Technology (ICRTIT'2014)*; Chennai. 2014 Apr 10-12. p. 1–6.
12. Handa A, Newcombe RA, Angeli A, Davison AJ. Applications of Legendre-Fenchel transformation to computer vision problems. Department of Computing at Imperial College London. DTR11-7 45; 2011.
13. Rudin LI, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*. 1992 Nov; 60(1–4):259–68.
14. Bioucas-Dias JM, Figueiredo MAT. Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing. 2nd Workshop on Evolution in Remote Sensing (WHISPERS) and Hyperspectral Image and Signal Processing; Reykjavik. 2010 Jun 14-16. p. 1–4.
15. Afonso MV, Bioucas-Dias JM, Figueiredo MAT. Fast image recovery using variable splitting and constrained optimization *IEEE Transactions on Image Processing*. 2010 Sep; 19(9):2345–56.