ELSEVIER
CrossMark
Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 93 (2016) 416 – 423
Procedia Computer Science

6th International Conference On Advances In Computing & Communications, ICACC 2016, 6-8 September 2016, Cochin, India

Application of Least Square Denoising to improve ADMM based Hyperspectral Image Classification

Srivatsa S[a,*], Aleena Ajay[a], Chandni C K[a], V Sowmya[a], Soman K P[a]

[a]*Centre for Computational Engineering and Networking (CEN)*
Amrita School of Engineering, Coimbatore
Amrita Vishwa Vidyapeetham, Amrita University, India

Abstract

Hyperspectral images contain a huge amount of spatial and spectral information so that, almost any type of Earth feature can be discriminated from any other feature. But, for this classification to be possible, it is to be ensured that there is as less noise as possible in the captured data. Unfortunately, noise is unavoidable in nature and most hyperspectral images need denoising before they can be processed for classification work. In this paper, we are presenting a new approach for denoising hyperspectral images based on Least Square Regularization. Then, the hyperspectral data is classified using Basis Pursuit classifier, a constrained L1 minimization problem. To improve the time requirement for classification, Alternating Direction Method of Multipliers (ADMM) solver is used instead of CVX (convex optimization) solver. The method proposed is compared with other existing denoising methods such as Legendre-Fenchel (LF), Wavelet thresholding and Total Variation (TV) . It is observed that the proposed Least Square (LS) denoising method improves classification accuracy much better than other existing denoising techniques . Even with fewer training sets, the proposed denoising technique yields better classification accuracy, thus proving least square denoising to be a powerful denoising technique.

Peer-review under responsibility of the Organizing Committee of ICACC 2016
Keywords: Denoising; hyperspectral image classification; basis pursuit; ADMM algorithm; Least Square; Legendre-Fenchel transformation.

1. Introduction

A new trend in remote sensing field is the Hyperspectral Imagery (HSI), which differs from other imaging systems in the sense that the number of bands captured by the imaging sensor is 100 to 200 or even greater. In a hyperspectral image, contiguous or non-contiguous bands of around 10nm bandwidth are available within the range of 400-2500 nm in the electromagnetic spectrum. The presence of a huge number of bands in a hyperspectral data makes it a hub of information resource, which can be used to classify features from captured image with more precision and detail. They can be used to precisely differentiate between land cover types and also to detect minerals, perform precision farming

* Corresponding author. Tel.: +91-9066656707.
E-mail address: srivatsa-s@live.com

Peer-review under responsibility of the Organizing Committee of ICACC 2016
doi:10.1016/j.procs.2016.07.228

and urban planning, etc.[1] In recent years, many algorithms are being developed for many hyperspectral applications namely, image classification, unmixing of bands, target detection, sub-pixel mapping, pansharpening, etc.

Classification of objects using HSI data is not contextual, where contextual means focusing on the relationship between nearby pixels. Normal image classification is performed by grouping pixels to represent land cover features such as urban, forest, agriculture, etc. But in HSI classification, each pixel vector is classified into different categories. Pixel vectors with similar characteristics are classified into the same group. A number of new classifiers are being developed and experimented on HSI data. Important classifiers used till date are Support Vector Machines (SVM)[2], Polynomial based Multinomial Logistic Regression (MLR), Minimum Spanning Forest (MSF), Orthogonal Matching Pursuit[3], etc. Some of the classifiers are sparsity-based, including the one we have employed in this work. Basis pursuit (BP), which is used here for classification, decomposes a signal to an optimal superposition of dictionary elements. The optimization criterion of BP is L1-norm of coefficients. It is superior in terms of stability and super-resolution over many other methods like Best Ortho Basis, Matching Pursuit and more[4]. The L1-norm problem is solved by using ADMM (Alternating Direction Method of Multipliers). It has been shown that ADMM is much faster than other solvers like CVX (convex optimization)[2].

But, the accuracy of classification is reduced by the presence of noise in the images. So, hyperspectral images need to be pre-processed before they are classified. Here, a new denoising technique is put forward based on Least Square (LS) Regularization which effectively denoises the HSI data and improves classification accuracy. Also, the proposed LS denoising is much faster than other denoising methods, thus reducing pre-processing time requirement. LS based denoising is compared with other denoising techniques such as Legendre-Fenchel (LF)[2], Wavelet based denoising and Total Variation (TV) denoising techniques. Standard hyperspectral datasets such as Indian pines, Pavia University and Salinas Scene[1,2,3] are used for experimental purposes. This paper explains the application of LS denoising to hyperspectral images in section 2, classification of hyperspectral images in section 3, and section 4 discusses the experimental results and analysis.

2. Least Square based Hyperspectral Image denoising

The objective of denoising is to obtain a clean signal x from a given noisy signal y. Each pixel may be represented by x_{ijb} for signal x and y_{ijb} for signal y. The indices i, j and b are the row position, column position and the band number respectively for the signals x and y. Assuming that the noise is additive zero-mean Gaussian, we can represent denoised signal y as

$$y_{ijb} = x_{ijb} + w_b \tag{1}$$

where, w_b is the noise component with a standard deviation of σ and is band-dependent. Solution to this problem is not unique as it is a problem of estimating $\hat{x}$ from y, which has an L2 fidelity term, $\hat{x} = argmin_x \|y - x\|_2^2$. This term is extended for denoising approach as a Least Square problem, which is formulated as

$$\min_x \|y - x\|_2^2 + \lambda \, \|\mathbf{D}x\|_2^2 \tag{2}$$

which states that the equation is to be minimized with respect to x. The terms λ and D are the control parameter and a second order difference matrix respectively. Minimizing the first term in equation (2) forces output to be similar to the original noisy signal. Minimization of second term leads to noise removal by smoothing the signal. So, λ is used to control the degree of smoothness and holding similarity to the noisy signal. Solving the Least Square problem leads to the following result

$$x = (\mathbf{I} + \lambda \mathbf{D}^T \mathbf{D})^{-1} y \tag{3}$$

For denoising 2D signal, i.e. an image, the LS solution is first applied on the columns of the 2D matrix, then applied on its rows. So, the procedure of denoising with Least Square is a matter of a simple matrix operation. It makes computation complexity to be reduced and requires lesser time to denoise when compared to techniques like Legendre-Fenchel and Wavelet denoising.

3. Hyperspectral Image Classification

In the supervised classification of hyperspectral images, a set of train pixels, usually 10% to 20% of hyperspectral data is chosen randomly from each of the class, as mentioned by the ground truth. Rest of the pixels or all the pixels are chosen for testing the classifier. Classifier generates class labels for each of the test pixels. In a sparsity-based algorithm, each class is represented in a lower dimensional subspace, leading to faster classification task. Classification is done by representing test samples in a sparse way with respect to the training samples. In this work, a sparsity-based classifier known as Basis Pursuit is used as the classifier.

3.1. Basis Pursuit

Chen and Donoho[4] have suggested a truly global optimization based decomposition technique called Basis Pursuit (BP). In this technique, a signal is decomposed into an optimal superposition of dictionary elements, using L1-norm optimization criterion. A dictionary D is a collection of signal waveforms $(\phi_\gamma)_{\gamma\in\Gamma}$, and the decomposition of signal s is

$$s = \sum_{\gamma\in\Gamma} \alpha_\gamma \phi_\gamma + R \tag{4}$$

where, R is a residue. This dictionary is an overcomplete dictionary. So, the decomposition is not unique and some elements in the dictionary may be represented in terms of other elements. It has the advantage of providing adaptation, allowing us to choose among many representations that suits best for our purpose. The main advantages of using Basis Pursuit classifier are its speed, sparsity, perfect separation, super-resolution and stability.

Out of many possible solutions to $\phi\alpha = s$, BP chooses a coefficient with minimum L1-norm: min $||\alpha||_1$ subject to $\phi\alpha = s$. When the signal is having noise above $\sigma > 0$, BP performs approximate decomposition by solving

$$min||\phi\alpha - s||_2^2 + \lambda_n||\alpha||_1 \tag{5}$$

where, $\lambda_n = \sigma\sqrt{2log(\#D)}$, $\#D$ indicates number of distinct vectors in the dictionary. This convex optimization problem is fed to ADMM solver. The obtained residues are used for classifying the pixels into different classes.

3.2. The Alternating Direction Method of Multipliers (ADMM) Algorithm

The ADMM algorithm is used to solve the convex optimization problems by breaking them into smaller problems and hence making them easier to handle. ADMM takes the fortune of superior convergence property of method of multipliers to blend with decomposability of dual ascent[5]. The algorithm solves the optimization problem as

$$\text{minimize } w(m) + r(n), \text{ subject to } Am + Bn = c \tag{6}$$

Here, w and r are assumed to be convex functions. The optimal output for this problem is denoted by

$$p^* = \inf \{w(m) + r(n)) \mid Am + Bn = c\} \tag{7}$$

The augmented Lagrangian may be expressed as

$$L_\rho(m, n, y) = w(m) + r(n) + y^T(Am + Bn - c) + (\rho/2)\,||Am + Bn - c||_2^2 \tag{8}$$

The algorithm consists of following iterations –

$$m^{k+1} := \text{argmin}_m\, L_\rho(m, n^k, y^k) \tag{9}$$

$$n^{k+1} := \text{argmin}_n\, L_\rho(m^{k+1}, n, y^k) \tag{10}$$

$$y^{k+1} := y^k + \rho(Am^{k+1} + Bn^{k+1} - c),\ \rho > 0 \tag{11}$$

The algorithm consists of minimization steps for m and n in equations (10) and (11), and a dual variable update in equation (12). The terms m and n are updated in an alternating fashion, hence name ‘alternating direction’[5]. This technique is an extended and advanced version of ‘method of multipliers’ algorithm.

Application of ADMM to solve Basis Pursuit is used to obtain residue for all the test vectors across all the classes. Then the text pixel is assigned to the class for which it has least residue value. Following this, a map that consists of class values for each pixel vector of original or denoised hyperspectral image is developed. This map is the classified output for the given hyperspectral data.

3.3. Mapping to Hyperspectral image classification

The classification task comprises of mainly two steps. One is to find the sparse solution m using basis pursuit which is solved by using ADMM. The other one is to calculate the minimum residue which corresponds to the class label for each of the test pixel vectors.

The 3-D hyperspectral image ($p \times q \times b$) is converted to 2-D ($b \times p * q$) by arranging all pixels in each spectral band as corresponding rows. Now, the rows corresponds to the spectral bands and the columns corresponds to the pixel vectors. A dictionary matrix A with size $b \times c$ is created which contains the training samples from every class. All the columns of the dictionary matrix are normalized. Sparse solution m is determined using basis pursuit solved using ADMM.The objective function of BP is

$$\hat{m} = \text{argmin } \|m\|_1, \text{ subject to } Am = t \tag{12}$$

where, $m \in R^c, A \in R^{b\times c}$ and $t \in R^b$.

The objective is to obtain a sparse solution. But, since the function is non-convex, convergence is difficult to obtain a solution. Hence, ADMM is utilized in order to speed up the convergence process[2]. The objective function of BP can be written in ADMM format as

$$\min g(m) + \|n\|_1, \text{ subject to } m - n = 0 \tag{13}$$

where g(m) is an indicator function of $m \in R^c$, given $Am = t$.

The update corresponding to m^{k+1}, n^{k+1} and y^{k+1} are obtained using ADMM. Here, y is the Lagrangian multiplier corresponding to the constraint $m - n = 0$. The augmented Lagrangian is given by,

$$L_\rho(m, n, y) = g(m) + \|n\|_1 + \rho/2\|m - n + y\|_1 \tag{14}$$

The update for m, n and y are given as,

$$m^{k+1} = \left(I - A^T(AA^T)^{-1}A\right)(n^k - y^k) + A^T(AA^T)^{-1}t \tag{15}$$

$$n^{k+1} = S_{1/\rho}(m^{k+1} + y^k) \tag{16}$$

where, $S_{1/\rho}$ is the soft thresholding operator.

$$y^{k+1} = y^k + m^{k+1} - n^{k+1} \tag{17}$$

Once the sparse solution m is obtained, the residue for test pixel vector is calculated using the equation

$$res_i = \|t - A^i_{new} * m\|_2 \tag{18}$$

where, i is iterated through no. of classes, t is the test pixel vector and A^i_{new} is the new dictionary matrix in which the columns, except those belongs to iteration value, are made zero. The test pixel vector is assigned to the class for which minimum residue is obtained.

4. Experimental Procedure

The intention of this work is to analyze the outcomes of the implementation of proposed LS based denoising technique before performing hyperspectral image classification. For this, we first classify the hyperspectral data without applying any denoising technique. Then denoising is applied to check improvement in classification. Denoising experiments are performed on standard datasets, namely Indian pines, Salinas Scene and Pavia University. SNR (Signal-to-Noise Ratio) values for each dataset after denoising using different methods are measured and recorded. Then the quality of denoising for each of the techniques is validated through classification experiment on Indian pines dataset. The accuracy of the classification is found out by comparing the classified data with the ground truth data and is recorded. Then, the different denoising methods, viz. total variation, Legendre-Fenchel, wavelet and proposed least square denoising are compared for their performance.

SNR calculation is a numerical approach for analyzing the amount of removal of noise from the signal or image. It is a measure of signal power to noise power, expressed in decibels (dB). An SNR calculation approach when there is no reference image present is explained by Linlin Xu in his paper[6]. In this approach, SNR for a signal is calculated as

$$SNR = 10log_{10}\frac{\Sigma_{ij}\hat{x}_{ijb}^2}{\Sigma_{ij}(\hat{x}_{ijb} - m_b)^2} \tag{19}$$

where, $\hat{x}_{ijb}$ is the denoised pixel, m_b is the mean value of $\{\hat{x}_{ijb}\}$ in an area where the pixels are homogeneous[6]. Estimation of SNR depends on the selection of homogeneous area. Ground truth data helps in finding the homogeneous area. SNR is calculated over all the bands and all the pixels in each band.

Classification of data provides the best way to analyze the denoising techniques. One of the problems in using SNR as ultimate quality measure tool is that an image which is more smooth will have higher SNR value although it might have lost edge information. So, classification helps in better understanding of the performance of the denoising techniques. Classification of hyperspectral data is performed after applying each of the denoising techniques. The classification accuracy is measured in each case and compared for analysis.

5. Results and Analysis

5.1. Denoised Quality Measurement

The hyperspectral data after being denoised are analyzed visually and numerically with the support of SNR measurement. The formula for measuring SNR of a denoised image is explained in the section (4). The figures 1, 2 and 3 show the denoised outputs of Indian pines, Pavia University and Salinas scene datasets respectively. It is observed that the proposed LS denoising method performs as good as existing techniques LF and wavelet denoising and is much better than TV denoising. It has been explained in section (2) that a control parameter λ controls the trade-off between signal smoothing and holds the similarity to original data. Increasing the value of λ improves noise removal but fades away the edge information. So, an optimal value of λ is to be chosen during denoising process. Table 1 shows the comparison of SNR values of different denoising methods on Indian Pines datasets. The values shown in the table are the average SNR obtained over all bands of the hyperspectral data.

Table 1. Comparison of SNR values for different denoising methods for Indian Pines data set

Denoising Technique	Average Signal to Noise Ratio (dB)		
	Indian Pines	Salinas	Pavia University
Original Noisy image	29.05	8.05	14.74
Total variation	32.47	10.33	16.31
Wavelet	29.96	8.37	15.48
Legendre-Fenchel	29.07	8.06	14.77
Least square	**29.82**	**8.55**	**15.41**

5.2. Classification Accuracy

Indian Pines hyperspectral image was captured by NASA's AVIRIS sensor. Indian Pines dataset constitutes of 145 $\times$ 145 pixels with 220 bands in the region 0.4 - 2.5 μm. It's ground truth consists of 16 classes. The assessment of

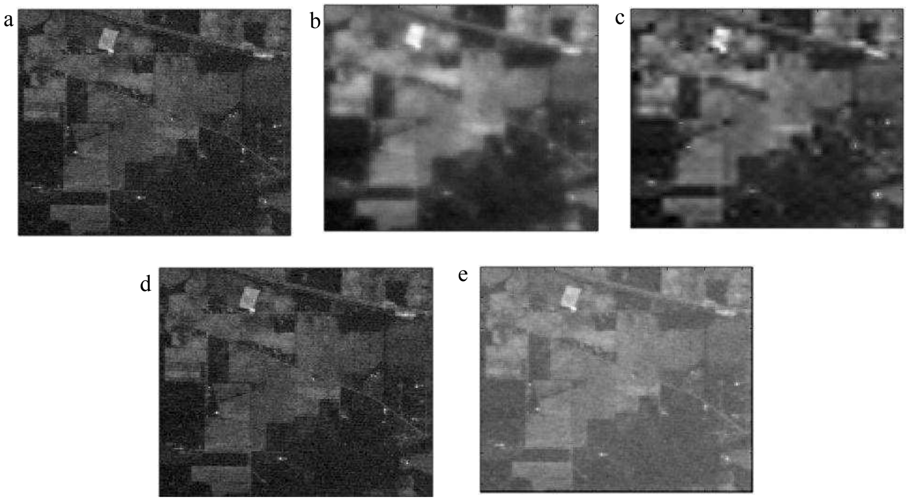

Fig. 1. Denoised outputs of band 3 of Indianpines dataset: (a). Noisy band; (b). TV denoised; (c). wavelet denoised; (d). LF denoised; (e). proposed LS denoised

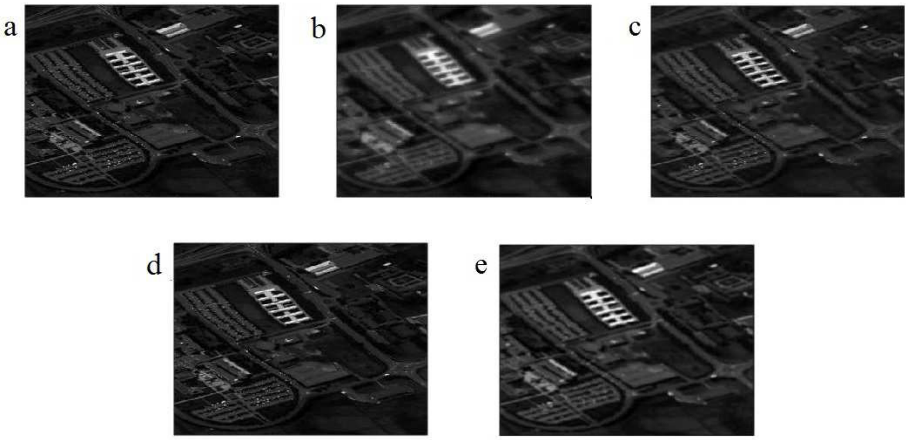

Fig. 2. Denoised outputs of band 15 of Pavia University dataset: (a). Noisy band; (b). TV denoised; (c). wavelet denoised; (d). LF denoised; (e). proposed LS denoised

classification is done both subjectively and objectively. Visual interpretation is used for subjective assessment whereas objective assessment is done through classification accuracy calculation. The important terms in objective assessment are

$$\text{classwise accuracy, } CA = \frac{\text{No. of pixels which are correctly classified in each class}}{\text{Total no. of pixels in each class}} \times 100 \tag{20}$$

$$\text{overall accuracy, } OA = \frac{\text{No. of pixels which are correctly classified}}{\text{Total no. of pixels}} \times 100 \tag{21}$$

$$\text{average accuracy, } AA = \frac{\sum_{c=1}^{n} CA_c}{n} \tag{22}$$

where, CA_c is the classwise accuracy of c^{th} class and n is the total number of classes in the hyperspectral image.

$$\text{Kappa coefficient} = \frac{P * C - S}{P^2 - S} \tag{23}$$

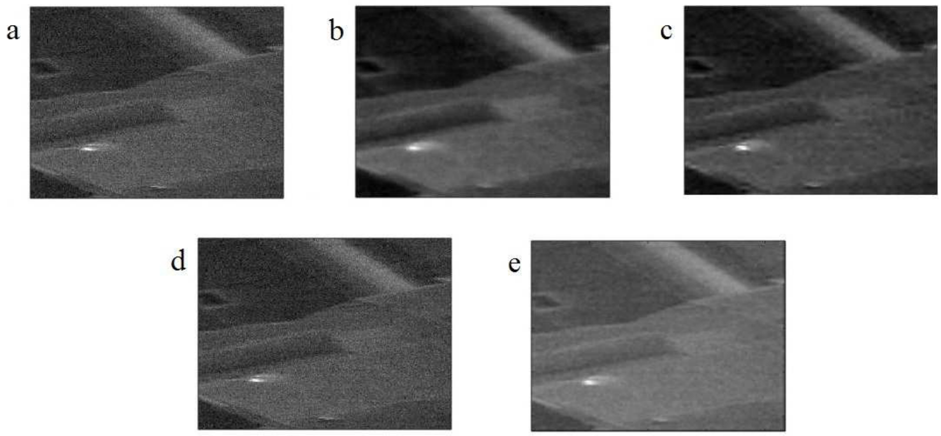

Fig. 3. Denoised outputs of band 2 of Salinas Scene dataset: (a). Noisy band; (b). TV denoised; (c). wavelet denoised; (d). LF denoised; (e). proposed LS denoised

where, P is the total number of pixels, C represents the number of pixels that are accurately classified and S is the sum of the product of rows and columns of the confusion matrix[2].

Table 2. Classification quality analysis for different denoising techniques on Indian Pines dataset

Denoising Technique	Overall Accuracy (OA%)	Average Accuracy (AA%)	Kappa Coefficient
Original Noisy image	69.158	58.8	0.6449
Total variation	97.6583	95.27	0.9733
Wavelet	90.8284	85.85	0.8952
Legendre-Fenchel	97.1802	95.99	0.9679
Least square	**98.2047**	**96.89**	**0.9795**

Table 2 contains the objective analysis of classification quality under each of the denoising methods. Figure 4 shows the original ground truth image of Indian pines dataset and classified outputs under different denoise techniques. It can be seen that the improvement in classification after denoising is obvious and better for the proposed least square technique than any other denoising techniques.

Improvement in classification for each case may be obtained by providing more training data. The outputs shown in the table 2 is for 10% training input. Increasing training input gives improved accuracy but consumes more time to execute the algorithm.

6. Conclusion

In this paper, we have proposed an extremely simple, yet effective denoising approach based on least square regression. The proposed LS technique has been verified for its capability with the help of SNR calculation and hyperspectral image classification. SNR calculations show that the proposed denoise technique is competitive with existing advanced denoised techniques like Legendre-Fenchel and wavelet-based methods. Classification of hyperspectral data has been carried out with Basis Pursuit classifier, which uses ADMM for solving the L1 optimization problem. The improvement in classification after denoising with the least square method is appreciable and is better than other denoising techniques discussed in this paper. So, the least square technique can emerge as an alternative to existing denoising techniques for hyperspectral data, which is the future of the remote sensing field.

Accurate measurement of SNR, when there is no reference image available is a difficult task and better solution needs to be introduced. The method implemented in this work is based on the selection of a homogeneous area in the

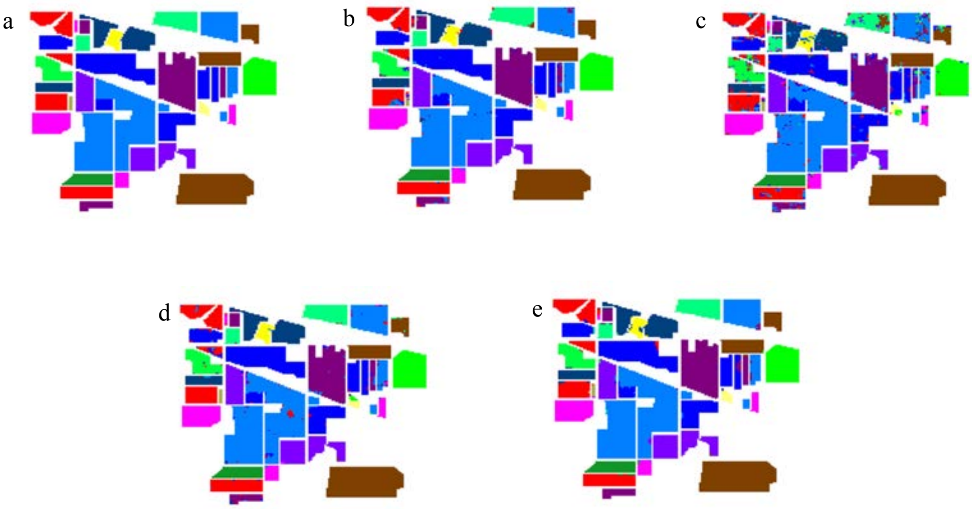

Fig. 4. Classified outputs under different denoising techniques of Indian pines dataset (a). Original ground truth; (b). Total variation [OA = 97.6583%]; (c). Wavelet [OA = 90.8284%]; (d). Legendre-Fenchel [OA = 97.1802%]; (e). Least square [OA = 98.2047%]

image. In our future work, we can consider the implementation of more advanced methods for SNR estimation. Also, the difference matrix, D, in the least square solution has the potential to improve denoising technique to much better levels. So, varying the values of the D matrix may help to improve denoise quality. These ideas will be considered as future work.

References

1. Yuan, Q., Zhang, L., Shen, H.. Hyperspectral image denoising employing a spectral–spatial adaptive total variation model. *Geoscience and Remote Sensing, IEEE Transactions on* 2012;**50**(10):3660–3677.
2. Aswathy, C., Sowmya, V., Soman, K.. Admm based hyperspectral image classification improved by denoising using legendre fenchel transformation. *Indian Journal of Science and Technology* 2015;**8**(24):1.
3. Zou, J., Li, W., Huang, X., Du, Q.. Classification of hyperspectral urban data using adaptive simultaneous orthogonal matching pursuit. *Journal of Applied Remote Sensing* 2014;**8**(1):085099–085099.
4. Chen, S., Donoho, D.. Basis pursuit 1994;**1**:41–44.
5. 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.
6. Xu, L., Li, F., Wong, A., Clausi, D.A.. Hyperspectral image denoising using a spatial–spectral monte carlo sampling approach. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of* 2015;**8**(6):3025–3038.