TWO-STAGE DENOISING METHOD FOR HYPERSPECTRAL IMAGES COMBINING KPCA AND TOTAL VARIATION

Wenzhi Liao, Jan Aelterman, Hiêp Quang Luong, Aleksandra Pižurica, Wilfried Philips

Ghent University-TELIN-IPI-iMinds, Sint-Pietersnieuwstraat 41, B-9000 Ghent, Belgium

ABSTRACT

This paper presents a two-stage denoising method for hyperspectral image (HSI) by combining kernel principal component analysis (KPCA) and total variation (TV). In the first stage, we use KPCA denoising to reduce spectrally uncorrelated noise. In the second stage, the information content is largely separated from the remaining noise by means of principal component analysis (PCA). The remaining noise is then efficiently removed by fast primal-dual TV denoising in low-energy PCA channels. Experimental results on simulated and real HSIs are very encouraging.

Index Terms— Hyperspectral images, denoising, kernel principal component analysis, total variation, classification.

1. INTRODUCTION

Despite advances in hyper-spectral sensor technology, image noise is unavoidably present, which can affect information retrieval and content interpretation. Using denoising as a pre-processing tool will improve various post-processing tasks, e.g. classification, target detection, unmixing, etc.

Many techniques were reported for noise reduction in hyperspectral data [1, 2, 3, 4, 5]. In [3], a multilinear algebra method was proposed to simultaneously reduce the spectral dimension and noise in hyperspectral images. Othman and Qian [2] proposed a hybrid spatial-spectral derivative-domain wavelet shrinkage noise reduction approach. The approach of [4] proposed a cubic total variation model for hyperspectral image denoising by combining the 2-D TV model for spatial domain and the 1-D TV model for spectral domain. The approach of [5] employed a spectral-spatial adaptive TV model for hyperspectral image denoising. Recently, Chen et al. [1] proposed a denoising algorithm for hyperspectral images with reasonably good SNR by using principal component analysis and wavelet shrinkage. The algorithm utilized PCA to decorrelate the fine features of the data cube from the noise, and then reduced the noise only in the noisy low-energy PCA output channels with wavelet shrinkage denoising.

However, when dealing with HSIs with low signal-to-noise ratio (SNR), the first k PCA channels also contain a

lot of noise. Therefore, this paper proposes a two-stage denoising method for HSIs. The proposed method first uses kernel PCA to reduce spectrally uncorrelated noise without sacrificing important information content. Then we use PCA to decorrelate the resulting data of the first stage from the remaining noise and employ a fast primal-dual total variation to reduce the noise only in the low-energy PCA output channels. The results demonstrate effectiveness of the proposed method and superior performance over recent methods like [1], both visually and quantitatively (in terms of PSNR and subsequent classification accuracy). The remainder of the paper is organized as follows. The proposed denoising method is detailed in Section 2. Section 3 presents the experimental results. Finally, conclusions are drawn in Section 4.

2. PROPOSED METHOD

We developed in this section a two-stage denoising method for HSI, which builds on some ideas from [1]. In the first stage, KPCA is used to reduce spectrally uncorrelated noise while preserving most information in HSI. The second stage of our approach is similar to [1], which utilizes the PCA to decorrelate the fine features of the data cube from the noise, reducing the noise only in the noisy low-energy PCA output channels. We use a first-order primal-dual total variation method [9] to remove the noise in the low-energy PCA channels, due to its computational efficiency and edge preserving property. The flow chart of the proposed two-stage denoising method for HSI is shown in Fig.1.

2.1. KPCA denoising

Kernel Principal Component Analysis [11] is a nonlinear version of PCA using techniques of kernel methods, which is more suitable to describe higher-order complex and nonlinear distributions. Its application to reducing the dimensionality of hyperspectral data sets has been investigated in [6, 7]. Suppose $\mathbf{x}_i \in R^B$ ($i = 1, 2, \cdots, n$) is the original data, where B is the spectral bands of HSI. There exists a function φ, which can map the original data into a higher or infinite dimensional Hilbert space. KPCA solves the eigenvectors by performing eigen decomposition on kernel space. A new data set can be obtained in
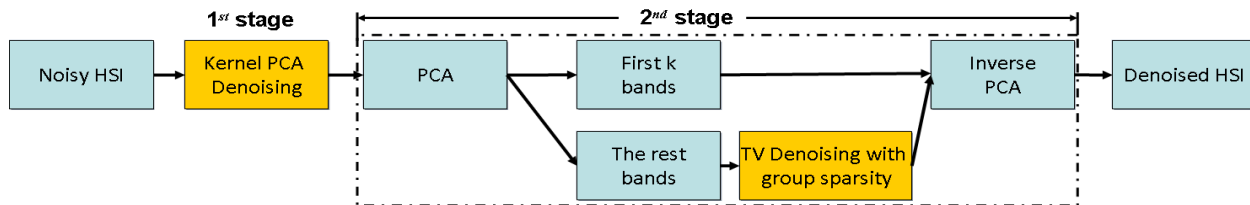
Fig. 1: Flow chart of proposed two-stage denoising method for HSI.

the feature space $\mathbf{\Phi} = (\varphi(\mathbf{x}_1), \varphi(\mathbf{x}_2), \cdots, \varphi(\mathbf{x}_n))$. φ is an implicit function, which cannot be calculated directly, but some kernel functions can be used by performing inner product between the two samples $\mathbf{x}_i$ and $\mathbf{x}_j$ in the original space, $\kappa_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j) = \varphi^T(\mathbf{x}_i)\varphi(\mathbf{x}_j)$. The most widely used Mercer kernels [11] include Gaussian kernel: $\kappa(\mathbf{x}_i, \mathbf{x}_j) = exp(-\frac{||\mathbf{x}_i - \mathbf{x}_j||^2}{2\delta^2})$.

Another interesting application of KPCA is in image denoising [8]. In order to denoise a given input data, KPCA first transforms the inputs into a higher dimensional feature space and then carries out linear PCA in the feature space to obtain the principal components. Denoising is then carried out by projecting the transformed pattern onto the most dominant eigenvectors (in our experiments, we just keep the most dominant eigenvectors, which represent 99% of the cumulative variance). By throwing away the rest of the very low-energy KPCA output channels, we can remove spectrally uncorrelated noise in HSI without losing too much information. In order to obtain the denoised pattern in the original input space (called pre-image [11]), a reverse transform is needed. However, an exact pre-image may not exist (i.e. the inverse transform of φ may not exist), as noted in [11] and [8]. Therefore, instead of obtaining the exact pre-image, one uses algorithms to estimate an approximating pre-image. For more details about KPCA denoising, we refer the interested reader to [8].

The number of extracted kernel principal components, which represent 99% of the cumulative variance depends on the number of total training samples and the parameters in the selected kernel function. In our experiments, 1000 samples were randomly selected to train and construct the training kernel matrix: Gaussian kernel function with $\delta = \frac{1}{n \times n}\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}\kappa_{ij}^2}$, where n is the total number of the training samples, $\kappa_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j)$ is the element of the kernel matrix.

2.2. TV denoising with group sparsity

KPCA suppresses well spectrally uncorrelated noise, while preserving most information of HSI. The noise in the first PCA output channels is suppressed quite well after KPCA denoising, as shown in Fig. 2. However, the resulting data still contain some remaining noise. Therefore, in the second
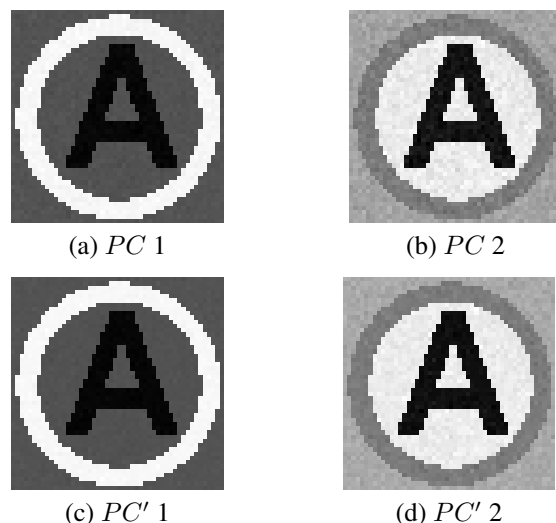

(a) $PC\ 1$ (b) $PC\ 2$

(c) $PC'\ 1$ (d) $PC'\ 2$

Fig. 2: The first two principal components before denoising (top row) and after KPCA denoising (bottom row).

stage, we employ a similar approach as in [1] for suppressing this remaining noise. We utilize the PCA to decorrelate the fine features of the data cube from the noise, reducing the noise only in the noisy low-energy PCA output channels. We use total variation denoising, which is widely used in image processing applications [4, 5, 9], because of its good edge preserving performance and also due to its computational efficiency, which is important when processing large data cubes. We propose to use group sparsity [12] to combine all the gradient coefficients (TV norm) over the different channels at the same spatial position into one group such that the resolving operator (soft-thresholding) does not act component-wise but treats the group as one vector.

Let u denote unknown noise-free data that we want to estimate, and f the resulting data from the low-energy PCA output channels after stage 1. The optimization problem that we consider is:

$$\hat{u} = \arg\min_{u}\sum_{i}^{B}||\nabla u||_2 + \frac{\lambda}{2}||f - u||_2^2 \quad (1)$$

where $\hat{u}$ is the estimate of the noise-free data u (the result of denoising), ∇ the discrete gradient operator. The first term on the right side of (1) is the TV group sparsity regularization, which is responsible for smoothing (i.e. noise reduction). The

Fig. 3: HSI data sets used in our experiments. Left: false color image of Indian Pines; middle: ground truth; right: simulated HSI using the average spectral reflection of same class in the ground truth.
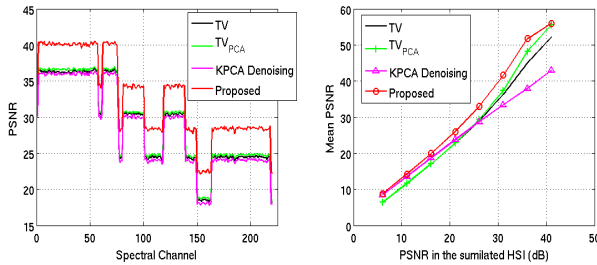

Fig. 4: Denoising results on the simulated HSI. Left: resulting PSNR per spectral band (PSNR of the input HSI is 25 dB). Right: mean PSNR with different PSNR in the simulated HSI after adding additive white Gaussian noise.

second term $||f - u||_2^2$ is commonly referred to data fidelity term. The parameter λ controls the relative contribution of the data fidelity term and smoothing term.

In our implementation, we used a very fast and memory efficient first-order primal-dual algorithm [9] for solving the optimization problem (1).

3. EXPERIMENTAL RESULTS

We used the AVIRIS Indian Pines image with 220 bands of size 145 lines by 145 samples, originally from Multispec$^©$. The whole scene contains 16 classes, ranging in size from 20 to 2468 pixels. 13 classes were selected for the experiments. The resulting false color composition and the groundtruth are shown in Fig. 3. We simulated a $50 \times 50 \times 220$ HSI with four classes in the groundtruth, the spectral reflection of the simulated HSI is the average of all data cube in the same selected class. By adding different amounts of noise in the simulated HSI, we compared peak SNR (PSNR) of each spectral band and the mean PSNR of all bands among the following methods: TV (primal-dual TV denoising on original HSI), TV$_{PCA}$ (first using PCA to decorrelate the HSI and then denoise only on the low-energy output PCA channels with TV), KPCA denoising and the proposed method.

The results in Fig. 4 show that KPCA denoising can reduce the noise in HSIs by removing 1% of low-energy output channels, and performs better than TV and TV$_{PCA}$ methods in case of HSIs with lower PSNR. The performances can be enhanced by removing more low-energy output channels, but this will lead to the loss of some image details. TV$_{PCA}$ produced better results than applying TV denoising directly on the original HSI. This is because of PCA can decorrelate the HSI. By using two-stage denoising, the proposed method outperforms the other methods.
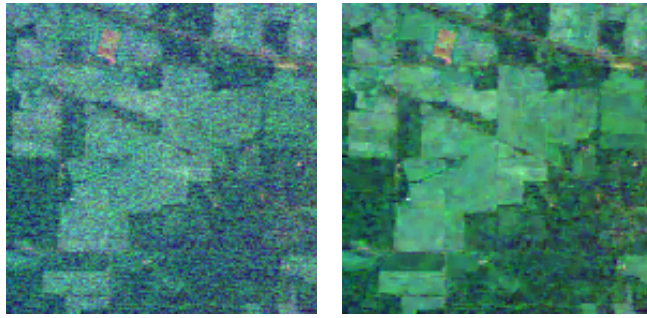
To further verify the effectiveness of our proposed denoising approach, we applied the denoising methods to the real HSI of AVIRIS Indian Pines, and then compared the classification results by using the HSI before and after denoising. In our application, we visually select the parameter λ in TV denoising, which can also be done by cross-validation [13]. 20 labeled samples per class of its ground truth are used to train the support vector machine (SVM) classifier [14]. Fig. 5 shows the denoised results with combination of three very noisy bands (bands 2, 109 and 219), Fig. 6 shows the classification results using SVM on the whole HSI.

The following conclusions can be drawn from the experiments:

- Applying TV denoising directly on the original HSI results in a noticeable loss of some detailed information, such as edges and texture. KPCA behaves much more conservative: spectrally uncorrelated noise is reduced and image details are well preserved. The use of PCA to decorrelate the HSI and denoising only the low-energy output PCA channels performs better than denoising directly on the original noisy HSI. The proposed two-stage method shows superiors performance in terms of noise reduction and detail preservation.

- The classification result has been greatly improved by the denoising process. The classification result of the original noisy HSI is rather poor. This is because of the effect of the strong noise information in most of the bands, and strong correlation between different spectral bands. However, the results were greatly improved by noise reduction, with smoother classification maps and higher overall classification accuracy (OA). Of the four classification results from the different denoising methods, it is shown that our proposed approach produces the best classification result with the highest OA of 83.5%.

4. CONCLUSION

A two-stage denoising method for hyperspectral images is proposed in this paper. We first use KPCA denoising to reduce spectrally uncorrelated noise while preserving most information in the HSI. Then we use PCA to decorrelate the resulting data from the first stage, and employ a fast primal-dual total variation denoising to remove noise only in the low-energy output PCA channels. The results on simulated and on real HSI show that the proposed method outperforms the tra-

(a)

(b)

(c)

(d)

(e)

Fig. 5: Denoising results on real hyperspectral image of AVIRIS Indian Pines. (a) Image composed of original noisy bands 2, 109 and 219, (b) TV, (c) TV$_{PCA}$, (d) KPCA denoising, and (e) the proposed two-stage denoising method.
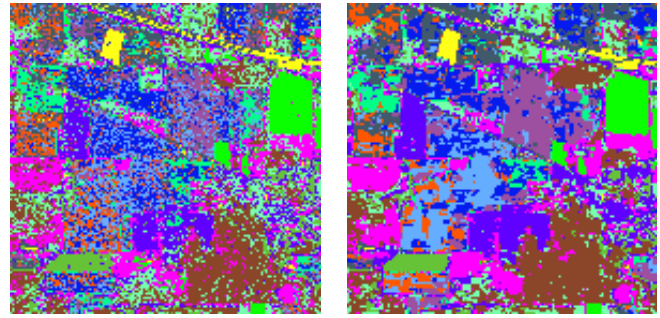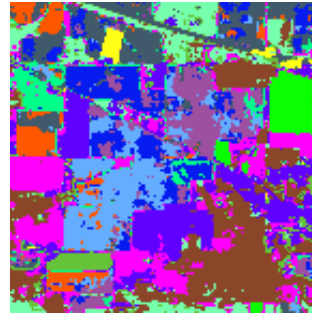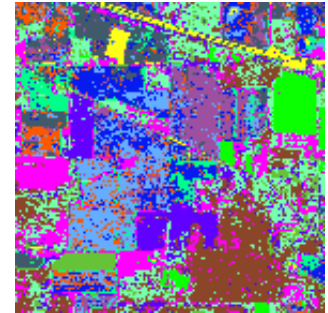

(a) $OA = 50.8\%$

(b) $OA = 64.4\%$

(c) $OA = 74.1\%$

(d) $OA = 63.8\%$

(e) $OA = 83.5\%$

Fig. 6: Classification results with SVM classifier applied on the original data or data after denoising with different methods, as indicated: (a) Original HSI, (b) TV, (c) TV$_{PCA}$, (d) KPCA denoising, and (e) proposed two-stage denoising method.

ditional one-stage denoising methods, and improves greatly the classification accuracy.

Acknowledgment

5. REFERENCES

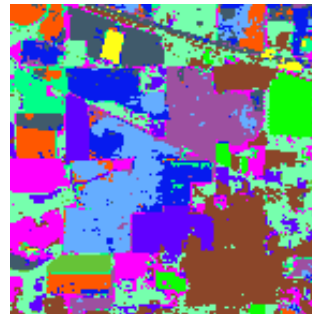[1] G. Chen, S. Qian, and W. Xie, "Denoising of Hyperspectral Imagery Using Principal Component Analysis and Wavelet Shrinkage," *IEEE Trans. Geosci. Remote Sens.*, 49(3):973–980, 2011.

[2] H. Othman and S. Qian, "Noise reduction of hyperspectral imagery using hybrid spatial-spectral derivative-domain wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, 44(2):397–408, 2006.

[3] D. Letexier and S. Bourennane, "Noise Removal from Hyperspectral Images by Multidimensional Filtering," *IEEE Trans. Geosci. Remote Sens.*, 46(7):2061–2069, 2008.

[4] H. Zhang, "Hyperspectral Image Denoising With Cubic Total Variation Model", *International Society of Photogrammetry and Remote Sensing Congress (ISPRS 2012)*, vol. 1-7, pp. 95–98, 2012.

[5] Q. Yuan, L. Zhang and H. Shen, "Hyperspectral Image Denoising Employing a Spectral-Spatial Adaptive Total Variation Model," *IEEE Trans. Geosci. Remote Sens.*, 50(10):3660–3677, 2012.

[6] M. Fauvel, J. Chanussot and J.A. Benediktsson, "Kernel principal component analysis for feature reduction in hyperspectrale images analysis," *In Proceedings of the 7th Signal Processing Symposium*, pp. 238–241, 2006.

[7] W. Liao, A. Pižurica, W. Philips and Y. Pi, "A Fast Iterative PCA Feature Extraction for Hyperspectral Images", *In Proceedings of 2010 IEEE International Conference on Image Processing (ICIP 2010)*, pp. 1317–1320, 2010.

[8] P.M. Rasmussen, T.J. Abrahamsen, K.H. Madsen, L.K. Hansen, "Nonlinear analysis of neuroimages with kernel principal component analysis and pre-image estimation," *Neuroimage*, 60(3):1807–1818, 2012.

[9] H. Luong, B. Goossens, J. Aelterman, A. Pižurica, W. Philips, "A Primal-Dual Algorithm For Joint Demosaicing And Deconvolution", *In Proceedings of 2012 IEEE International Conference on Image Processing (ICIP 2012)*, pp. 2801–2804, 2012.

[10] US Army Topographic Engineering Center, HyperCube, *http://www.tec.army.mil/Hypercube/*.

[11] B. Scholkopf, A.J. Smola and K.R. Muller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, pp. 1299–1319, 1998.

[12] A. Majumdar and R.K. Ward, "Compressed sensing of color images," *Signal Processing*, vol. 90, pp. 3122–3127, 2010.

[13] S. Arlot, A. Celisse, "A survey of cross-validation procedures for model selection", *Statistics Surveys*, vol. 4, no. 0, pp. 40–79, 2010.

[14] C.C. Chang and C.J. Lin, "LIBSVM: a library for support vector machines," *http://www.csie.ntu.edu.tw/ cjlin/libsvm*, 2001.