Wavelet packets for multi- and hyper-spectral imagery

J. J. Benedetto, W. Czaja, M. Ehler, C. Flake, M. Hirn

J. J. Benedetto, W. Czaja, M. Ehler, C. Flake, M. Hirn, "Wavelet packets for multi- and hyper-spectral imagery," Proc. SPIE 7535, Wavelet Applications in Industrial Processing VII, 753508 (4 February 2010); doi: 10.1117/12.843039

Event: IS&T/SPIE Electronic Imaging, 2010, San Jose, California, United States

Wavelet Packets for multi- and hyper-spectral imagery

J. J. Benedetto, W. Czaja, M. Ehler, C. Flake, M. Hirn

Norbert Wiener Center
Department of Mathematics
University of Maryland
College Park, MD 20742

Abstract

State of the art dimension reduction and classification schemes in multi- and hyper-spectral imaging rely primarily on the information contained in the spectral component. To better capture the joint spatial and spectral data distribution we combine the Wavelet Packet Transform with the linear dimension reduction method of Principal Component Analysis. Each spectral band is decomposed by means of the Wavelet Packet Transform and we consider a joint entropy across all the spectral bands as a tool to exploit the spatial information. Dimension reduction is then applied to the Wavelet Packets coefficients. We present examples of this technique for hyper-spectral satellite imaging. We also investigate the role of various shrinkage techniques to model non-linearity in our approach.

1 Introduction

Multi-spectral techniques for data analysis have been successfully applied to detect and classify objects in areas ranging from human pathology to geophysics and to satellite imaging. A classical assumption about existence of a low dimensional representation of the high dimensional multi-spectral data leads to a special role played by dimension reduction. Principal Component Analysis (PCA) is one common example of dimension reduction, [16]. It is a linear transform that determines the 'directions' in the data by maximizing the captured variance. If the data manifold is linear, then PCA is the optimal choice.

Known dimension reduction and classification methods in multi- and hyper-spectral imaging use mostly the spectral component of the information and do not exploit the knowledge of the spatial distribution. Our aim in this paper is to present a model in

Wavelet Applications in Industrial Processing VII, edited by Frédéric Truchetet, Olivier Laligant,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 7535, 753508 · © 2010 SPIE-IS&T
CCC code: 0277-786X/10/$18 · doi: 10.1117/12.843039

which both, spatial and spectral, components of the data play a significant role. We shall achieve this by utilizing a multiresolution analysis approach. It is well known that wavelets and their multilevel decompositions are well-suited for the analysis of spatial characteristics. We shall combine wavelet analysis with PCA to capture spatial and spectral data distributions for multi-spectral imagery. One approach to achieve this is to use the sequential wavelet decomposition for each spectral pixel vector, respectively, cf. [14], forming a multi-spectral dataset of coefficients, and then to perform PCA on the coefficients. This, however, does not effectively utilize the spatial information provided inherently in each band. On the other hand, while decomposing the multi-spectral datasets band-wise, one must find an optimal way to simultaneously incorporate spatial information with multiple spectra.

The Discrete Wavelet Transform (DWT) is an iterative scheme that splits the signal into approximation and detail coefficients, cf. [13] for its application to hyper-spectral data. Each level is computed by passing through only the previous approximation coefficients. The Wavelet Packet Transform (WPT), however, decomposes both the approximation and the detail coefficients. Contrary to the DWT, this yields a full wavelet tree decomposition that offers flexibility in the choice of reconstruction coefficients. The best basis algorithm by Coifman and Wickerhauser finds the optimal subtree for reconstruction, i.e., the best coefficient set with respect to a given entropy measure, [8]. See Figure 1 for a visualization of the subtree concept.

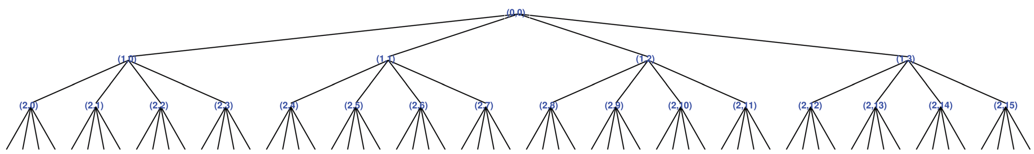
Wavelet Packets and the entropy concept offer the flexibility to transfer spatial information from one band to the other. We separately decompose each band by using the WPT, but we consider a joint entropy over all bands. The joint entropy still respects the correlation between different frequency bands in the wavelet domain. The optimal subtree is determined by evaluating the joint entropy. Each node of the common subtree then is a set of coefficient vectors whose entries refer to different spectral bands. PCA is applied to these vectors and we then reconstruct the original coefficients by applying the inverse WPT. Fusing spectral and spatial information significantly improves the output of classification schemes.

The outline of our presentation is as follows: In Section 2, we present the Wavelet Packet Transform adapted to the multi-spectral setting. We specify the concept of joint entropy in Section 2.1. In Section 2.2, we present a number of appropriate shrinkage strategies that have already been considered in the literature in related image processing contexts. In Section 2.3 we discuss the flexibility of our approach. We apply our method to hyper-spectral satellite images in Section 3.

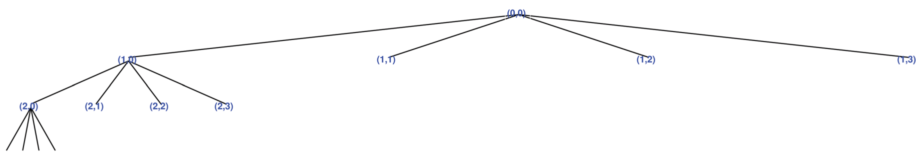
2 The Wavelet Packets Approach

2.1 Joint Entropy

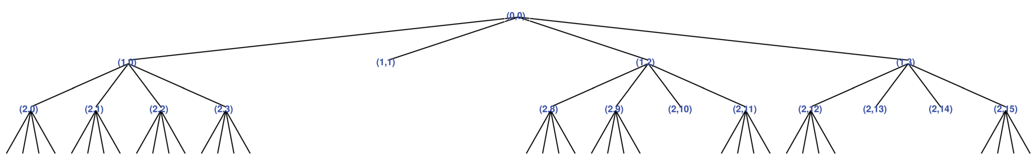
Given a full wavelet coefficients tree, let $\mathcal{T}$ denote the collection of all subtrees. The entropy E (or cost function) of a Wavelet Packet decomposition is a nonnegative map $E : \mathcal{T} \to \mathbb{R}$. Given n spectral bands, we now decompose each band by means of the WPT. This yields a separate entropy E_i, $i = 1, \dots, n$, for each band, respectively. We



(a) full wavelet packet tree



(b) wavelet tree



(c) possible wavelet packet subtree

Figure 1: The entropy determines the subtree that is chosen for reconstruction

define the joint entropy $\mathcal{E}$ through the weighted ℓ_p norm:

$$\mathcal{E} = \sum_{i=1}^{n} w_i |E_i|^p, \tag{1}$$

where $0 < p \leq 2$ and the weight vector $\{w_i\}_{i=1}^n$ is to be specified. Minimizing the entropy yields the best subtree for reconstruction.

If the assumption that the underlying data manifold in the wavelet domain is linear does not hold (i.e., in the case of a nonlinear regime) then PCA cannot capture the essential information by itself. There are two ways to handle the nonlinearity. We could replace PCA with a nonlinear dimension reduction such as Locally Linear Embedding (LLE) [17], Laplacian Eigenmaps (LE) [1, 2], Hessian LLE [10], and Diffusion Wavelets/Diffusion maps [4, 5, 6, 7]. However, these methods are known to be computationally very expensive. Our simpler approach is to apply shrinkage to the wavelet coefficients – the classical concept for compression and denoising, see the following Section 2.2 for few specific shrinkage rules. In the context of multi- and hyper-spectral imagery, shrinkage concentrates spatial and spectral information into few coefficients in the wavelet domain. Finally, we apply PCA to the shrinked coefficients and then transform back.

Remark 2.1. Through shrinkage, the output of the WPT becomes a nonlinear approximation of the signal. Moreover, shrinkage means smoothening of the underlying manifold. This is a step towards linearization.

2.2 Shrinkage Rules

Decomposing signals into simple building blocks and reconstructing from shrinked coefficients are widely used in signal representation and processing, e.g., wavelet shrinkage is applied to noise and clutter reduction in speckled SAR images, cf. [15]. Many different shrinkage strategies have been proposed in literature. Let us recall few common shrinkage rules. We restrict us to real coefficients $x \in \mathbb{R}$, see also Figure 2. Soft- and hard-shrinkage are by far the most widely used strategies and they can be based on a common theoretical fundament that involves the Besov norm of the signal, see [3, 9]. *Soft-shrinkage* is given by

$$\varrho_s(x, \alpha) = (x - \text{sign}(x)\alpha)\mathbf{1}_{\{|x|>\alpha\}},$$

where the wavelet coefficient, that is to be shrinked, will replace x and the amount of shrinkage is α. While soft-shrinkage still modifies large x, *hard-shrinkage*

$$\varrho_h(x, \alpha) = x\mathbf{1}_{\{|x|>\alpha\}}$$

keeps large coefficients, but has a discontinuity. The *nonnegative garotte-shrinkage* rule

$$\varrho_g(x, \alpha) = (x - \frac{\alpha^2}{x})\mathbf{1}_{\{|x|>\alpha\}}$$

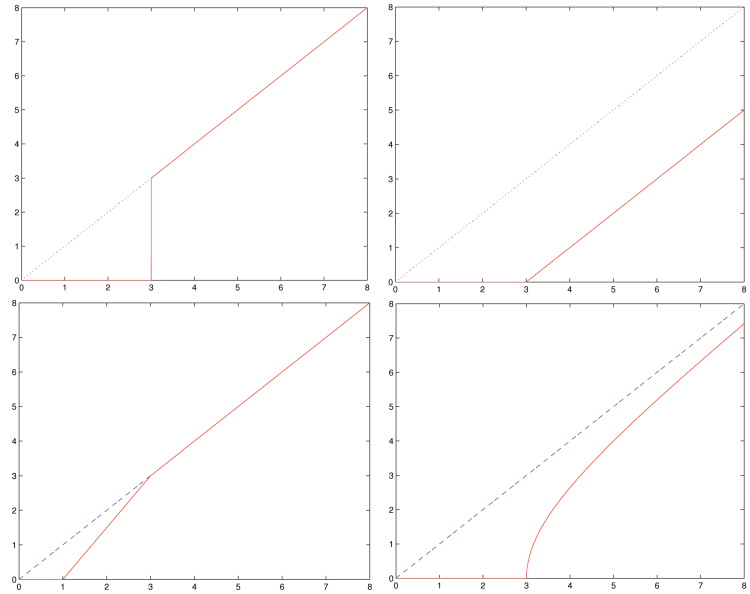


Figure 2: The shrinkage rules hard-, soft-, firm with $\beta = 1$, and nonnegative garotte

is continuous and large coefficients are left almost unaltered. It has been successfully applied to image denoising in [11]. Bruce and Gao proposed *firm-shrinkage*

$$\varrho_{f,\beta}(x, \alpha) = x\mathbf{1}_{\{|x|>\beta\}} + \text{sign}(x)\frac{\beta(|x| - \alpha)}{\beta - \alpha}\mathbf{1}_{\{\alpha \leq |x| \leq \beta\}}$$

in [12] as a piecewise linear method.

2.3 Flexibility of the Proposed Method

Our proposed method provides significant flexibility to fit any particular dataset or field of application: The entropy E_i for each band has to be chosen. MATLAB already provides the different entropy functions *shannon, log energy, threshold, sure, norm* within the wavelet toolbox. There is also the option to apply a user defined entropy function. The weights $\{\omega_i\}_i^n$ and the parameter p also can be specified. For smaller values of p, the sparsity is emphasized within the joint entropy concept. Finally, we have to pick a specific shrinkage rule ϱ and must determine the amount of shrinkage α. There is a large variety of different shrinkage strategies in the literature and we have presented only four popular shrinkage rules in the previous Section 2.2. In the following Section 3, it turns out that the shrinkage rule can significantly influence the classification results.

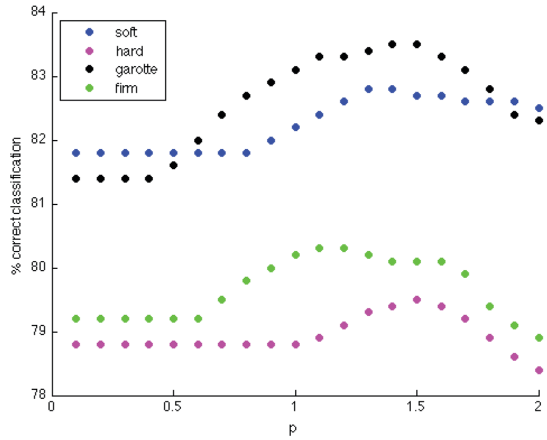


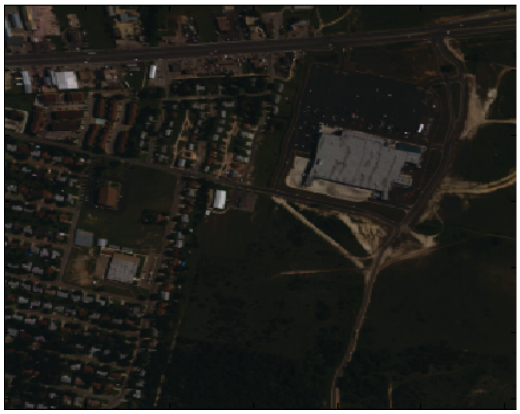
Figure 3: Classification results for different shrinkage rules, varying $0 < p \leq 2$, and fixed α with 30 principal components

3 Numerical Results for Hyper-spectral Satellite Images

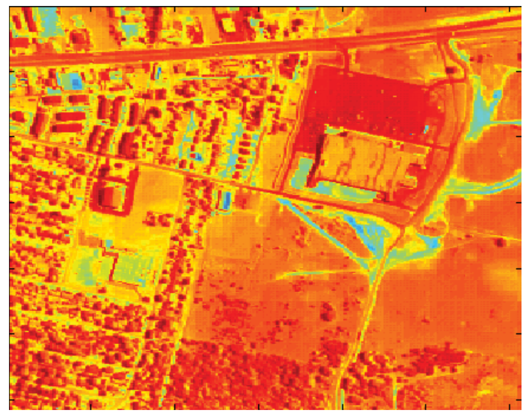
The present section is not only dedicated to the application of our proposed method but also to shed some light on the choice of p and the shrinkage rule. For simplicity, we restrict our considerations to the Haar-wavelet, Shannon entropy, and the constant weights $\omega_i = 1$, for all $i = 1, \ldots, n$.

We have applied the proposed method to the 'Urban' dataset, a collection of 161 spectral bands of 307×307 pixels. Since we have groundtruth, we can compare our proposed method with standard PCA on the original dataset with respect to classification. We use 15 to 50 principal components that capture more than 99% of the variance. To choose p, we have run extensive experiments with varying shrinkage parameters α, see Figure 3 for classification results with fixed α and fixed number of principal components. These experiments suggest $p \approx 1.3$ for soft-shrinkage, which also seems to be an appropriate choice for other shrinkage rules. A classification result for soft-shrinkage and $p = 1.3$ is given in Figure 4.

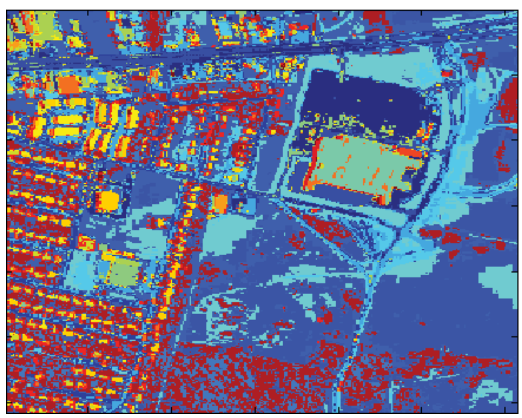
For fixed $p = 1.3$ and fixed shrinkage parameter α, we have compared the performance of the shrinkage rules presented in Section 2.2. It turns out that applying PCA to shrinked Wavelet Packet coefficients outperforms simple PCA that is directly applied to the dataset, see Figure 5. Nonnegative garotte is the best choice under these conditions and it shows the same trend as soft-shrinkage with varying number of principal components. It should be mentioned that the results are hard to distinguish on a visual basis. We have chosen soft-shrinkage for Figure 6, but the other shrinkage rules produce classes that are visually very similar. This suggests that the primary differences between shrinkage rules occur at the spatial edges between one class and the other.



(a) RGB image of Urban dataset



(b) reconstruction from the first PCA component of the wavelet domain



(c) vector angle classification using 15 PCA components from shrinked wavelet packet coefficients

Figure 4: Classification Results on Urban dataset for soft-shrinkage and $p = 1.3$

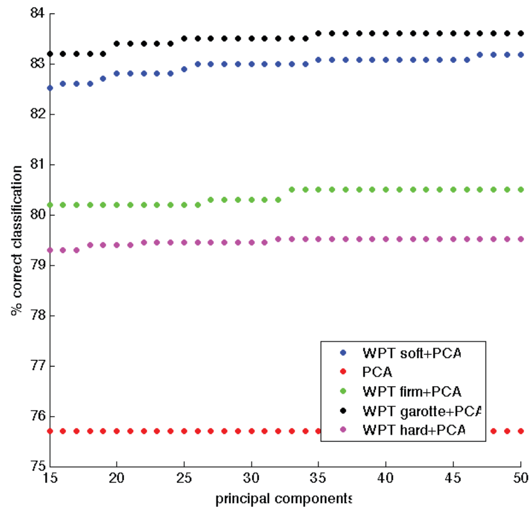


Figure 5: The combination of WPT with PCA outperforms simple PCA

4 Conclusion

We have proposed a new method to analyze multi- and hyper-spectral imagery which exploits spatial and spectral contents of the data. Using the spatial information about the image together with the spectral contents, rather than relying on purely spectral information, significantly improves classification results. More specifically, we have introduced a joint entropy function for the Wavelet Packets and verified its usefulness in multi- and hyper-spectral imagery as a tool that combines spectral and spatial information. To account for the non-linearity of the dataset, we have investigated a number of non-linear shrinkage techniques. We then applied PCA and the inverse wavelet transform. Our experiments suggest that, within this scheme, the non-negative garotte shrinkage and a choice of the entropy parameter $p \approx 1.3$ provide optimal results among our proposed shrinkage rules and the parameter range $0 < p \leq 2$.

5 Future Directions

More extensive experiments and an elaborate analysis are required to find the optimal choice of the entropies E_i, the shrinkage parameter α, and the weights $\{\omega_i\}_{i=1}^n$. The use of redundant wavelet systems, tight or bi-frames, may offer more flexibility to adapt the wavelet system to the specific characteristics of the multi-spectral data set. We plan to address these topics in a forthcoming paper.

Despite their computational costs, it seems natural to use nonlinear dimension reduction techniques that replace linear PCA. Preliminary experiments, however, have shown that further adaptation is required when applying them to the wavelet domain.

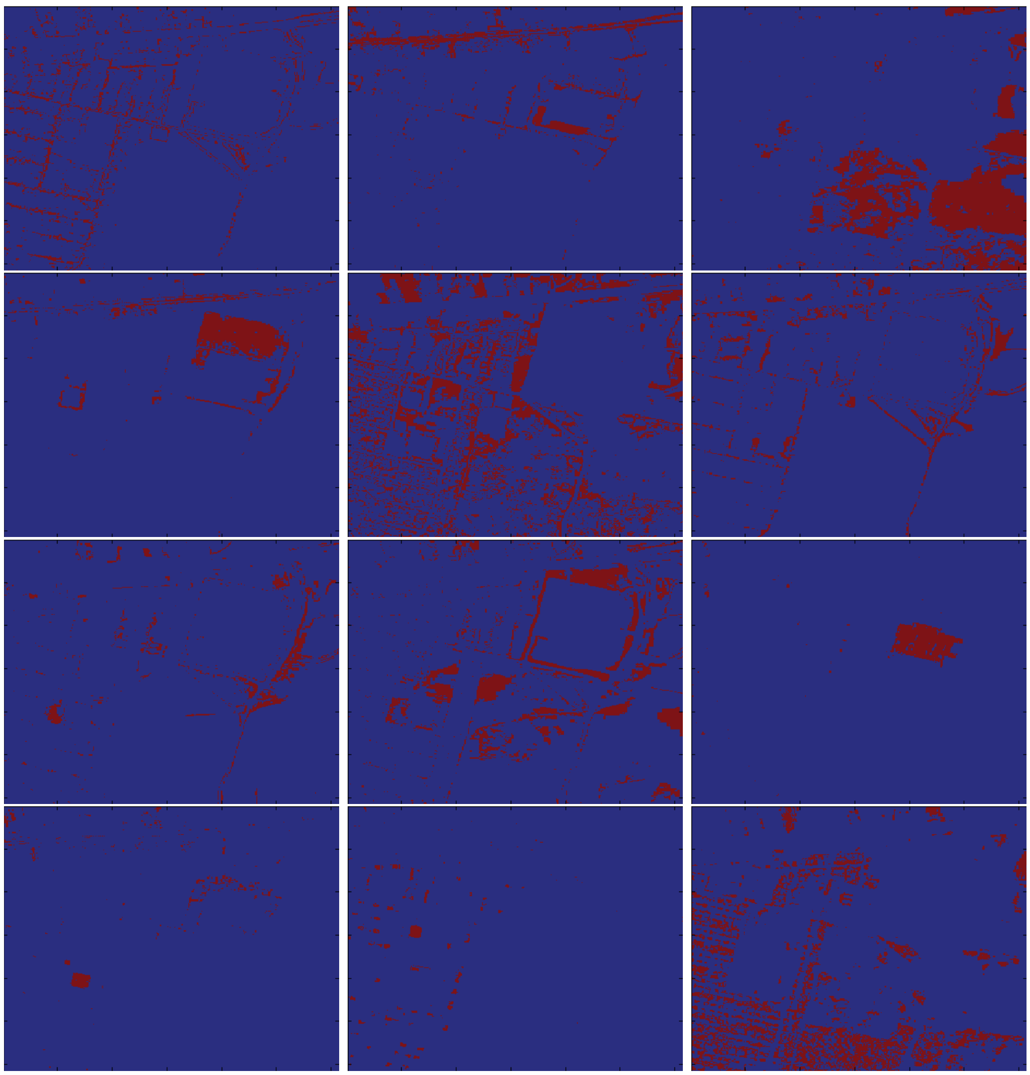


Figure 6: classes of the vector angle classification scheme for the Urban dataset

Acknowledgments

This material is based in part upon work supported by National Geospatial-Intelligence Agency through grant HM15820810009 and by Office of Naval Research under contract N000140910144.

References

[1] M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. *NIPS*, 14, 2002.

[2] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural. Comput.*, 15(6):1373–1396, 2003.

[3] A. Chambolle, R. A. DeVore, N. Y. Lee, and B. J. Lucier. Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Trans. Image Process.*, 7:319–335, 1998.

[4] R. R. Coifman and S. Lafon. Diffusion maps. *Appl. Comput. Harmon. Anal.*, 21(1):5–30, 2006.

[5] R. R. Coifman and S. Lafon. Geometric harmonics: A novel tool for multiscale out-of-sample extension of empirical functions. *Appl. Comput. Harmon. Anal.*, 21(1):31–52, 2006.

[6] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. J. Warner, and S. W. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data. part i: Diffusion maps. *Proc. Nat. Acad. Sci.*, 102:7426–7431, 2005.

[7] R. R. Coifman and M. Maggioni. Diffusion wavelets. *Appl. Comput. Harmon. Anal.*, 21(1):53–94, 2006.

[8] R. R. Coifman and M. V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory*, 38(2):713–718, 1992.

[9] D. Donoho and I. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81:425–455, 1994.

[10] D. L. Donoho and C. Grimes. Hessian eigenmaps: new locally linear embedding techniques for high-dimensional data. *Proc. of National Academy of Sciences*, 100:5591–5596, 2003.

[11] H. Y. Gao. Wavelet shrinkage denoising using the non-negative garotte. *J. Comput. Graph. Statist.*, 7(4):469–488, 1998.

[12] H. Y. Gao and Andrew G. Bruce. Waveshrink with firm shrinkage. *Statistica Sinica*, 7:855–874, 1997.

[13] L. Jiang, X. Chen, and G. Ni. Wavelet threshold denoising with four families of mother wavelets for hyperspectral data. *Proceedings of SPIE - The International Society for Optical Engineering*, 7156(71564D), 2009.

[14] L. Jiang, X. Chen, G. Ni, and S. Ge. Wavelet threshold denoising for hyperspectral data in spectral domain. *SPIE (ICEODPA)*, 7285, 2008.

[15] A. Niedermeier, E. Romaneessen, and S. Lehner. Detection of coastlines in SAR images using wavelet methods. *IEEE Trans. Geosci. Remote Sensing*, 38(5):2270–2281, 2000.

[16] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(7-12):559–572, Jul-Dec 1901.

[17] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.