A Robust High-dimensional Data Reduction Method

Longcun Jin[1], Wanggen Wan[1], Yongliang Wu[1], Bin Cui[1] Xiaoqing Yu[1] and Youyong Wu[1, 2]

[1]*School of Communication and Information Engineering, Shanghai University, Shanghai, China*
[2]*School of Information and Electronic Engineering, East China University of Technology, Fuzhou, Jiangxi, China*

Abstract—**In this paper, we propose a robust high-dimensional data reduction method. The model assumes that the pixel reflectance results from linear combinations of pure component spectra contaminated by an additive noise. The abundance parameters appearing in this model satisfy positivity and additive constraints. These constraints are naturally expressed in a Bayesian literature by using appropriate abundance prior distributions. The posterior distributions of the unknown model parameters are then derived. The proposed algorithm consists of Bayesian inductive cognition part and hierarchical reduction algorithm model part. The proposed reduction algorithm based on Bayesian inductive cognitive model is used to decide which dimensions are advantageous and to output the recommended dimensions of the hyperspectral image. The algorithm can be interpreted as a robust reduction inference method for a Bayesian inductive cognitive model. Experimental results on high-dimensional data demonstrate useful properties of the proposed reduction algorithm.**

Index Terms—**High-dimensional data, Bayesian model, Inductive cognitive, Hyperspectral image.**

I. INTRODUCTION

The reduction of hyperspectral image has been widely used in remote sensing signal processing for data analysis [1-4]. Its underlying assumption is based on the fact that all data sample vectors are mixed by a number of so-called endmembers assumed to be presented in the data. By virtue of this assumption, two models have been investigated in the past to model how mixing activities take place. One is the macrospectral mixture that describes a mixed pixel as a linear mixture endmembers opposed to the other model suggested by Hapke, referred to as intimate mixture that models a mixed pixel as a nonlinear mixture [5-6]. Another is how to estimate these endmembers once the number of endmembers is determined. Such statistics are said to be sufficient if they capture all the "relevant information" in the sample about the identity of reduction of hyperspectral image [7-8].

In this work we propose a way of quantifying this information using information theoretic notions and show how features which maximize this information can be extracted. Due to its link with the statistical concept of reduction, an view of this setup is a generalization of the problem of nonlinear regression. There is often more information between X and Y than can be captured by a single conditional expectation value [9-13]. Our problem then is to find several such functions, or regressions, that together capture more of the information and structure of the variables. As will be shown in this paper, this problem can be cast as finding dimension reduction of hyperspectral image which captures the mutual information in a two-way contingency table. It is thus related to a long line of work in statistics.

As explained above, the linear reduction of model is classically used to model the spectrum of a pixel in the observed scene. And the dimension reduction of hyperspectral image has already received much attention in the literature [3-4, 14-17]. Consequently, their estimation requires to use a quadratic programming algorithm with linear equalities and inequalities as constraints. Different estimators were developed using these ideas [18-20]. This paper studies a Bayesian estimator of Inductive Cognition which allows one to estimate the abundances in an nonlinear reduction of model. The proposed algorithm defines appropriate prior distributions for the unknown signal parameters and estimates these unknown parameters from their posterior distributions.

The prior distributions used in the present paper depend on hyperparameters which have to be determined. The unknown parameters in the posterior distributions requires to use appropriate simulation methods such as hierarchical Bayesian methods [21-23]. It is well known that the dimensionality of input space strongly affects performance of many classification methods. This requires the careful design of new algorithms that are able to handle hundreds of such spectral images at the same time minimizing the effects from the "curse of dimensionality". We first illustrate a well-known phenomenon in hyperspectral data. To address the characteristic that certain parts of the spectrum will provide a much richer descriptor for classification than other parts, some approaches such as a straightforward feature selection or a block-based approximation to the covariance matrix can be applied [24-25].

The remainder of the paper is organized as follows. An overview of the related work is given in Section Ⅱ. The proposed algorithm is discussed in Section Ⅲ. Section Ⅳ presents the experimental settings and performance evaluation. Section Ⅴ concludes this paper.

II. RELATED WORK

The natural variability of the material spectra and the noise added by the transmission media and sensor system make necessary the use of statistical methods for information extraction

and pattern recognition on hyperspectral data. Hyperspectral imaging technology has found application beyond earth remote sensing in agriculture, medicine, biology, pharmaceuticals, forensics, color vision, target detection, archaeology, and many others near field applications. However, classification of hyperspectral data is primarily made on a pixel by pixel basis with classification accuracy figures in the range 79%~84%, and they have not changed significantly in recent decade [4, 6, 10, 26-28]. The scale-space framework introduced by the diffusion equation has been also used for image reduction, in conjunction with level sets to detect movement in image sequences, information extraction and image restoration, registration, and classification integrating level sets in a common framework [17-18]. Continuous transformation of the original image into a space of progressively smoother images identified by the scale or level of image smoothing, in terms of pixel resolution.

In recent years, hyperspectral data reduction and classification are generally considered as a processing step for spectra classification, target detection or segmentation. For each of these problems, different methods exist. Methods of feature extraction like projection pursuit or Bayesian construct a lower dimension feature space by transforming the original data to preserve the most discriminable information content but change the physical meaning of the components [10, 13, 15, 20]. Spectral reduction methods, as discussed in [4, 13], reduce the number of spectra by keeping the representative ones. Such methods do not care about the spatial structure of hyperspectral data; hyperspectral data is only considered as a set of spectra. In [3], the authors use independent component analysis for band selection reduction method. The aim in this case is to identify the spectral signature that identifies these materials, each one having specific characteristics on certain bands. [24] utilized a Gaussian mixture model to analysis to reduce the size of the spectral vector. In our study, we account for the spatial structure of the data through a hierarchical Bayesian model about from that we also account for the noise and thus estimate a typical mean spectra for each class. The authors in [7] show that the reduction matrix obtained might be far from the true one. The classical hierarchical Bayesian model does not take into account any information about noise or any prior information on the signal of interest [29-31]. In this paper, we proposed a Bayesian estimation framework for the sources with a common Bayesian classification variable which is modeled as a inductive cognition field. Each value of this Bayesian variable corresponds to a characteristic mean spectrum of a given region.

III.　　THE PROPOSED ALGORITHM

Since Bayesian model with concentration hyperparameter α defines a prior on all partitions of the n_k hyperspectral image point n D_k (the value of α is directly related to the expected number of classes), the prior on the merged hypothesis is the relative mass of all n_k points belonging to one class versus all the other partitions of those n_k virtual reality data points consistent with the tree structure. The proposed algorithm is based Bayesian inductive cognitive model by an approximate infe-

rence approach. This can be computed bottom-up as the tree built shown in Fig. 1.

initialize each leaf i to have $d_i = \alpha$, $\pi_i = 1$
for each internal node $\mathbf{k}$ do
$$d_k = \alpha\Gamma(n_k) + d_{left_k}d_{right_k}$$
$$\pi_k = \frac{\alpha\Gamma(n_k)}{d_k}$$
end for

Fig. 1. To compute prior on merging, where $right_k$ ($left_k$) indexes the right (left) subtree of T_k and d_{right_k} (d_{left_k}) is the value of d computed for the right (left) child of internal nodes k.

The main idea in our Bayesian of inductive cognition model is to use the hyperspectral model (3) and the prior distributions (4) and

$$\pi(A) \propto N(A_0,\ \Gamma_0) \tag{1}$$

$$\pi(\Sigma_\varepsilon^{-1} \mid v_0,\ \Sigma_0) = |\Sigma_\varepsilon|^{\left(\frac{v_0 - m - 1}{2}\right)} \times \exp\left[-\frac{1}{2}Tr\left(\Sigma_\varepsilon^{-1}\Sigma_0^{-1}\right)\right] \tag{2}$$

$$\pi(\mu_{jk}) = N(\mu_0,\ v_0)$$
$$\pi(\sigma_{jk}^2) = IG(\alpha_0,\ \gamma_0) \tag{3}$$

to obtain the posterior law

$$p(s,\ z,\ \theta \mid x) \propto p(x \mid s,\ z,\ \theta)p(s,\ z \mid \theta)p(\theta) \tag{4}$$

where $\theta = (\theta_1,\ \theta_2)$ and $\theta_1 = (A, \Sigma_\varepsilon)$, $\theta_2 = (\{m_{jc},\ \sigma_{jk}^2\})$ represent the mixing matrix A and all the hyperparameters of the prior modeling of the noise and the sources. Here, the unknowns are first splitted in two categories: the Bayesian variables $(s,\ z)$ and all the hyperparameters θ, and then use the following iterative algorithm. Estimate $(s,\ z)$ using $p(s,\ z \mid \hat\theta,\ x)$ by $\hat s \sim p(s \mid \hat z, \hat\theta,\ x)$. In this paper, $\sim$ represents either $\arg\max$ or compute the MFA. To implement this algorithm, we need the following expressions: $p(s \mid z,\ \theta,\ x) \propto p(x \mid s,\ \theta_1)p(s \mid z,\ \theta_2)$. It is then easy to see that $p(s \mid z,\ \theta,\ x)$ is separable in r. $s(r)$ is estimated by its mean value $\bar s(r)$

$$p(s \mid z,\ \theta,\ x) = \prod_r p(s(r) \mid z(r),\ \theta,\ x(r)) = \prod_r N(\bar s(r),\ Bb(r))$$

with

$$\begin{cases} B(r) = \left[A^t\Sigma_\varepsilon^{-1}A + \Sigma_{z(r)}^{-1}\right]^{-1} \\ \bar s(r) = B(r)\left[A^t\Sigma_\varepsilon^{-1}x(r) + \Sigma_{z(r)}^{-1}m_{z(r)}\right] \end{cases} \tag{5}$$

In this relation, $m_{z(r)}$ is a vector of size K with all elements equal to zero except the kth element where $k = z(r)$ and $\Sigma_{z(r)}$ is a diagonal matrix of size $K \times K$ with all elements equal to zero except the kth diagonal where $k = z(r)$. Note that the prior law on z is

$$p(z \mid \beta) \propto \prod_r p(z(r) \mid z(r'),\ r' \in V(r),\ \beta) \tag{6}$$

$$\propto \exp\left[\sum_r \sum_{r' \in V(r)} \beta z'(r)z(r')\right] \tag{7}$$

The expression of the posterior $p(z|\theta, x')$ is

$$P(z|\theta, x') \propto \prod_r p(x(r)|z'(r), \theta) p(z'(r)|z'(r), r' \in V(r), \beta)$$

$$\propto \prod_r p(z'(r)|x(r), z'(r'), r' \in V(r), \theta, \beta)$$

In this paper, our Bayesian of reductive cognitive model consists then in approximation $p(z|\theta, x')$ by

$$q(z|\theta, x') \simeq \prod_r q(z'(r)|x(r), \bar{z}'(r'), r' \in V(r), \beta) \qquad (8)$$

where $\bar{z}'(r)$ represents the mean value of $z'(r)$, and

$$q(z'(r)|x(r), \bar{z}'(r'), r' \in V(r), \theta, \beta)$$

$$\propto p(x(r)|z'(r), \theta) p(z'(r)|\bar{z}'(r'), r' \in V(r), \beta)$$

$$\propto \exp\left[\beta\left(z'^t(r)\Sigma_r \bar{z}'(r')\right) + \ln p(x(r)|z'(r), \theta)\right]$$

and $\bar{z}'(r)$ is given by

$$\bar{z}'(r) = \frac{\sum_{z'(r)} z'(r) q(z'(r)|x(r), \bar{z}'(r'), r' \in V(r), \theta, \beta)}{\sum_{z'(r)} q(z'(r)|x(r), \bar{z}'(r'), r' \in V(r), \theta, \beta)} \qquad (9)$$

Our main goal in our proposed algorithm is to incorporate the idea of multi-scale hyperspectral data reduction into extended Bayesian of inductive cognitive transformations. However, if the reduction patterns do not have regular properties across the hyperspectral data, an adaptative scheme is needed to ensure the wonderful experimental result. In order to extend our algorithm model and closing operations to hyperspectral image, let us consider a hyperspectral image f defined on R^N. Given a Bayesian model of minimal size, extended opening by reconstruction is defined by

$$(f \circ B)^k(x, y) = \bigcup_{k \geq 1}\left[\delta_B^k(f \circ B|f)\right](x, y) \qquad (10)$$

where

$$\left[\delta_B^k(f \circ B|f)\right](x, y) = \left[\overbrace{\delta_B\delta_B\cdots\delta_B}^{k}(f \circ B|f)\right](x, y) \qquad (11)$$

The elementary term $\left[\delta_B(f \circ B|f)\right](x, y)$ is an extended geodesic dilation, defined as the maximum of the elementary dilation of the elementary dilation of $f \circ B$ using B at pixel (x, y) and the value of $f(x, y)$,

$$\left[\delta_B(f \circ B|f)\right](x, y) = \left\{\left[(f \circ B) \oplus B\right](x, y)\right\} \vee \left\{f(x, y)\right\} \qquad (12)$$

The elementary term $\left[\delta_B(f \circ B|f)\right](x, y)$ is an extended geodesic erosion, defined as the minimum of the elementary erosion of $f \bullet B$ using B at pixel (x, y) and the value of $f(x, y)$,

$$\left[\delta_B(f \bullet B|f)\right](x, y) = \left\{\left[(f \bullet B) \oplus B\right](x, y)\right\} \wedge \left\{f(x, y)\right\} \qquad (13)$$

Using Eqs. (18) and (20), extended morphological Bayesian models are created as follows. Let the vector $p_k^0(x, y)$ be the extended opening by reconstruction Bayesian model at the point (x, y) of the hyperspectral image f, defined by

$$p_k^0(x, y) = \left\{(f \circ B)^\lambda(x, y)\right\}, \quad \lambda = \{0, 1, \cdots, k\} \qquad (14)$$

And let $p_k^0(x, y)$ be the extended closing by reconstruction Bayesian model at the point (x, y) of the hyperspectral image f, defined by

$$p_k^\bullet(x, y) = \left\{(f \bullet B)^\lambda(x, y)\right\}, \quad \lambda = \{0, 1, \cdots, k\} \qquad (15)$$

Here $(f \bullet B)^0(x, y) = f(x, y) = (f \circ B)^0(x, y)$ for $\lambda = 0$ by the definition of extended opening and closing by Bayesian reconstruction. We define the derivative of the extended opening Bayesian model $\Delta p_k^0(x, y)$ as the vector

$$\Delta p_k^0(x, y) = \left\{Dist\left[(f \circ B)^\lambda(x, y), (f \circ B)^{\lambda-1}(x, y)\right]\right\}, \quad \lambda = \{0, 1, \cdots, k\} \qquad (16)$$

By duality, the derivative of the closing Bayesian model $\Delta p_k^\bullet(x, y)$ is the vector

$$\Delta p_k^\bullet(x, y) = \left\{Dist\left[(f \bullet B)^\lambda(x, y), (f \bullet B)^{\lambda-1}(x, y)\right]\right\}, \quad \lambda = \{0, 1, \cdots, k\} \qquad (17)$$

Given all of the above, the multi-scale opening characteristic $\Phi_k^0(x, y)$ at the point (x, y) of the image f is defined as the Bayesian model with the greatest associated value in $\Delta p_k^0(x, y)$

$$\Phi_k^0(x, y) = \lambda \in \{0, 1, \cdots, k\},$$

$$Dist\left[(f \circ B)^\lambda(x, y), (f \circ B)^{\lambda-1}(x, y)\right] = \bigcup \Delta p_k^0(x, y)$$

Similarly, the Bayesian model of multi-scale closing characteristic $\Phi_k^\bullet(x, y)$ is defined as the Bayesian model with greatest associated value in the derivative of the extended closing profile $\Delta p_k^\bullet(x, y)$:

$$\Phi_k^\bullet(x, y) = \lambda \in \{0, 1, \cdots, k\}$$

$$Dist\left[(f \bullet B)^\lambda(x, y), (f \bullet B)^{\lambda-1}(x, y)\right] = \bigcup \Delta p_k^\bullet(x, y)$$

The procedure architecture of the proposed reduction algorithm is shown in Fig. 2.
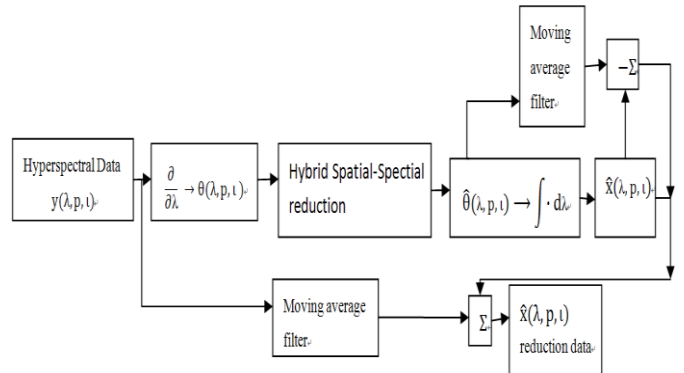

Fig. 2. Architecture of the reduction algorithm.

IV. EXPERIMENTAL RESULTS

In this part, we will present the effectiveness of the dimensional reduction approach based on Bayesian model, over the hyperspectral image data set. However, since the data is separated into various discontinuous intervals, the dimensional reduction approach cannot be applied as is and need to be adapted for the hyperspectral data. Thus, the regular Bayesian reduction algorithm can be applied as it is, on each interval. The decomposed hypespectral data output is composed of the decomposed output of each considered interval, appended to each other. This mechanism is illustrated in Fig. 4. The hyperspectral data interval decomposition can be done during the pre-processing stage.

Such algorithm is composed of three major phases: pre-processing, processing and post-processing. The

pre-processing initializes the problem variables and orders the intervals from the hyperspectral data file. The processing step determines for each pixel and its interval the maximum number of Bayesian dimensional reduction which leads to a viable correlation. The granularity of considered intervals used during the Bayesian processing can vary. It can be matching the number of intervals of the hyperspectral data, namely 17 intervals. On the other hand, the granularity can be coarser, to match the three major subintervals. This is possible since intervals composing those subintervals are generating contiguous spectrum ranges. However, a special care must be taken to conserve spectral contiguousness inside each major subinterval, i.e. reading intervals 1, 2, 3, 4 in this order, according to Table 1, for first major subinterval.

TABLE 1: INTERVALS FOR HYPERSPECTRAL SPECTROMETER LEVEL 1B DATA.

| Interval #	Spectral Cuton (µm)	Spectral Cutoff (µm)	# bands	Starting Channel	Ending Channel
1	3.7364	3.9169	118	0	117
2	4.11	4.3291	130	118	247
3	3.9149	4.11	116	248	363
4	4.3271	4.6085	150	364	513
5	6.9356	7.4769	192	514	705
6	6.2003	6.4934	104	706	809
7	6.5504	6.85	106	810	915
8	7.4745	7.7921	94	916	1009
9	7.8605	8.22	106	1010	1115
10	8.8073	9.4796	159	1116	1274
11	9.565	10.275	167	1275	1441
12	10.275	10.985	167	1442	1608
13	11.0704	11.7512	161	1609	1769
14	11.7431	12.685	167	1770	1936
15	12.7989	13.7457	167	1937	2103
16	13.7377	14.5533	144	2104	2247
17	14.6672	15.4	130	2248	2377

Table 1 shows the level of Bayesian dimension reduction achieved for different correlation values, as well as the resulting number of bands, per major sub-interval. Each sub-interval is identified by its interval components. The total reduction line gives the overall reduction efficiency. Table 1 is considering only three sub-intervals, concatenating close intervals to produce contiguous ranges in frequency domain. On the other hand, Table 2 is considering each interval independently. It can be noticed that by breaking the spectrum into smaller domains, a different level of reduction can be achieved for each interval.

In our work, we conduct some experiments in hyperspectral data dimensionality reduction in order to demonstrate the feasibility of the proposed algorithm. Our datacube was acquired using an Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). The original scene with size of 256×256 pixels, was acquired by the AVIRIS sensor, which is a mixed river/city area in south of china, early in the growing season. The scene comprises 220 spectral channels to be used in the experiments. These laboratory spectra, which were convolved in accordance

with AVIRIS wavelength specifications, will be used to assess endmenber signature purity in this work. At this point, it is important to note that many of the mineral spectra in the USGS library are not from the south of china area. Thus, the best match between hyperspectral image endmember and one in the USGS library spectra does not necessarily mean the best match the endmember. In addition, some minerals do not occur in pure form in the area, specifically at the 20-m spatial resolution of the sensor. For illustrative purposes, Fig. 3 shows high-dimension data expressed by computer graphics and a reduction result of hyperspectral data using our proposed reduction based on Bayesian inductive cognition model. It is useful that the results are reduced by our proposed algorithm.
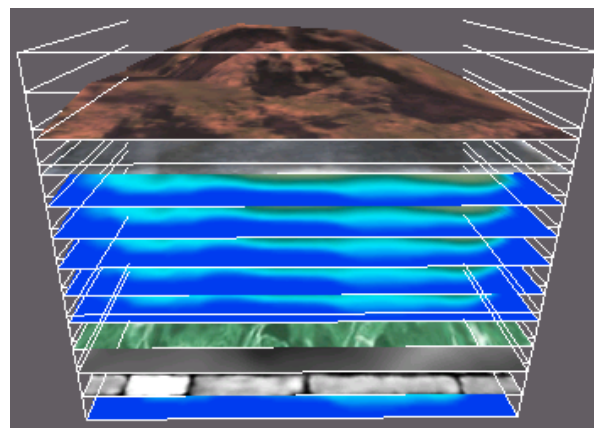
(a)
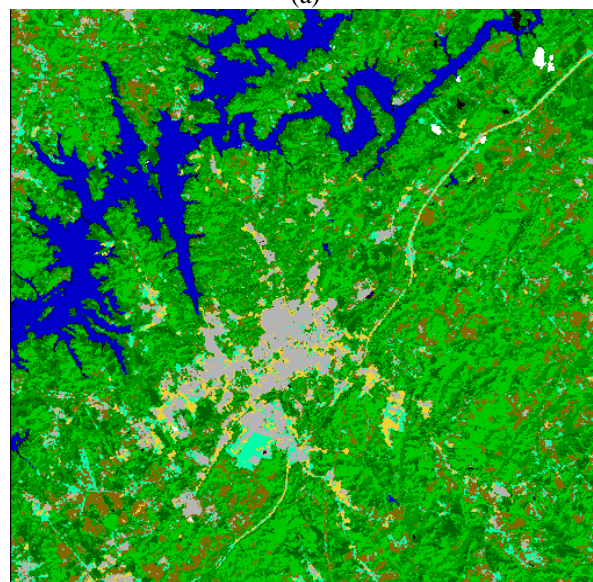
(b)

Fig. 3. (a)high-dimension data expressed by computer graphics (b)The Reduction Result of Hyperspectral Image. (Color Plate 11,12)

V.   CONCLUSION

This paper studied Bayesian inductive cognitive model for nonlinear hyperspectral reduction. The relationships between the different hyperspectral spectra were naturally expressed in a Bayesian context by the prior distributions adopted for the proposed model and their parameters. The posterior distribution of the unknown parameters related to the proposed model were

estimated by a Bayesian sampling strategy. These posterior distributions provided estimates of the unknown parameters but also information about their uncertainties such as standard deviations of confidence intervals. The proposed algorithms were developed depending whether the endmembers belonging to the mixture are known or belong to a known area. Simulation results conducted on real images illustrated the performance of the proposed reduction algorithm based on Bayesian inductive cognitive model. The hierarchical reduction methodologies developed in this paper could be modified to handle more complicated models. For instance, it would be interesting to extend the proposed algorithm to reduce hyperspectral data composed of homogenous regions surrounded by hyperspectral image by introducing dimensions correlation via Bayesian models.

Future work will investigate the use of complex hyperspectral data for the selection of the number of reduced bands and of the correlation threshold, as well as for detecting data anomalies. The dimension reduction technique based on Bayesian model will also be compared to the PCA technique to estimate the efficiency of each with the hyperspectral data.

ACKNOWLEDGEMENT

REFERENCES

[1] J. B. Tenenbaum, V. de Silva, J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Scince*, vol. 290, no. 5500, pp. 2319-2323, Dec. 2000.

[2] Antonio Plaza, Pablo Martinez, Javier Plaza, Rosa Perez. Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations. *IEEE Trans. on Geoscience and Remote Sensing*, vol. 43, no. 3, March. 2005.

[3] Jing Wang, Chein-I Chang. Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis. *IEEE Trans. on Geoscience and Remote Sensing*, vol. 44, no. 6, Jun. 2006.

[4] Hisham Othman, Shen-En Qian. Noise reduction of hyperspectral imagery using hybrid spatial-spectral derivative-domain wavelet shrinkage. *IEEE Trans. on Geoscience and Remote Sensing*, vol. 44, no. 2, Feb. 2006.

[5] Nicolas Dobigeon, Jean-Yves Tourneret, Chein- Ⅰ Chang. Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery. *IEEE Trans. on Signal Processing*, vol. 56, no. 7, Jul. 2008.

[6] Martin H. C. Law, Anil K. Jain. Incremental nonlinear dimensionality reduction by manifold learning. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, no. 3, March. 2006.

[7] Anish Mohan, Guillermo Sapiro, Edward Bosch. Spatially coherent nonlinear dimensionality reduction and segmentation of hyperspectral images. *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 2, April. 2007.

[8] Osnat Keren. Reduction of the average path length in binary decision diagrams by spectral methods. *IEEE Trans. on Computers*, vol. 57, no. 4, pp. 520-531, Apr. 2008.

[9] Cook R. Dennis. Fisher lecture: dimension reduction in regression. *Statistical Science*, vol. 22, no.1, pp. 1-26. 2007.

[10] Yanchang Zhao, Shichao Zhang. Generalized dimension-reduction framework for recent-biased time series analysis. *IEEE Trans. on Knowledge and Data Engineering*, vol. 18, no. 2, pp. 231-244, Feb. 2006.

[11] Jie Chen, Yousef Saad. Lanczos vectors versus singular vectors for effective dimension reduction. *IEEE Trans. on Knowledge and Data Engineering*, published online 5 Nov. 2008.

[12] Aboulnasr Hassanien, Sergiy A. Vorobyov. A robust adaptive dimension reduction technique with application to array processing. *IEEE Signal Processing Letters*, vol. 16, no. 1, Jan. 2009.

[13] Hisham Othman, Shen-En Qian. Noise reduction of hyperspectral imagery using hybrid spatial-spectral derivative-domain wavelet shrinkage. *IEEE Trans. On Geoscience and Remote Sensing*, vol. 44, no. 2, Feb. 2006.

[14] Jieping Ye, Ravi Janardan, Qi Li, Haesun Park. Feature reduction via generalized uncorrelated linear discriminant analysis. *IEEE Trans. on Knowledge and Data Engineering,* vol. 18, no. 10, Oct. 2006.

[15] Stephane Lafon, Ann B. Lee. Diffusion maps and coarse-graining: a unified framework for dimensionality reduction, graph partitioning, and data set parameterization. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, no. 9, Sep. 2006.

[16] Hua-Liang Wei, Stephen A. Billings. Feature subset selection and ranking for data dimensionality reduction. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 2009, no. 1, Jan. 2007.

[17] Huiwen Zeng, H. J. Trussell. Constrained dimensionality reduction using a mixed-norm penalty function with neural networks. *IEEE Trans. on Knowledge and Data Engineering*, First Published Online: 2009-4-24.

[18] Nicolas Dobigeon, Jean-Yves Tourneret, Chein- Ⅰ Chang. Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery. I*EEE Trans. on Signal Processing*, vol. 56, no. 7, pp. 2684-2695, Jul. 2008.

[19] Joshua B. Tenenbaum, Thomas L. Griffiths, Charles Kemp. Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, vol. 10, no. 7, pp. 309-318, Jul. 2008.

[20] Robert S. Lynch, Peter K. Willet. Bayesian classification and feature reduction using uniform Dirichlet priors. *IEEE Trans. on Systems, Man, and Cybernetics*, Part B, vol. 33, no. 3, pp. 448-464, Jun. 2003.

[21] Guangyi Chen, Shen-En Qian. Simultaneous dimensionality reduction and denoising of hyperspectral imagery using bivariate wavelet shrinking and principal component analysis. *Canadian Journal of Remote Sensing*, vol. 34, no. 5, pp. 337-454, Oct. 2008.

[22] Liangpei Zhang, Yanfei Zhong, Bo Huang, Jianya Gong, Pingxiang Li. Dimensionality reduction based on clonal selection for hyperspectral imagery. *IEEE Trans. on Geoscience and Remote Sensing*, vol. 45, no. 12, Dec. 2007.

[23] Are C. Jensen. Anne Schistad Solberg. Fast hyperspectral feature reduction using piecewise constant function approximations. *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 4, pp. 547-551, Oct. 2007.

[24] Jason K. Johnson, Alan S. Willsky. A recursive model-reduction method for approximate inference in Gaussian Markov random fields. *IEEE Trans. on Image Processing*, vol. 17, no. pp. 70-83, Jan. 2008.

[25] Xiaogang Wang, Xiaoxu Ma, W. Eric L. Grimson. Unsupervised activity perception in crowded and complicated scenes using hierarchical Bayesian models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 31, no. 3, pp. 539-555, Mar. 2009.

[26] Hongjun Zhou, Shigeyuki Sakane. Sensor planning for mobile robot localization: a hierarchical approach using a Bayesian net work and a particle filter. *IEEE Trans. on Robotics*, vol. 24, no. 2, pp. 481-488, Apr. 2008.

[27] Jin-Hyuk Hong, Youn-Suk Song, Sung-Bae Cho. Mixed –initiative human-robot interaction using hierarchical Bayesian networks. *IEEE Trans. on Systems, Man, and cybernetics—Part A: Systems and Humans*, vol. 37. no. 6, pp. 1158-1164, Nov. 2007.

[28] Manuel Davy, Christian Doncarli, Jean-Yves Tourneret. Classification of chirp signals using hierarchical Bayesian learning and MCMC methods. *IEEE Trans. on Signal Processing*, vol. 50, no. 2. pp. 377-388, Feb. 2002.

[29] Yanni Kouskoulas, Fawwaz T. Ulaby, Leland E. Pierce. The Bayesian hierarchical classifier (BHC) and its application to short vegetation using

multifrequency polarimetric SAR. *IEEE Trans. on Geoscience and Remote Sensing*, vol. 42, no. 2. pp. 469-477, Feb. 2004.

[30] Sanghoon Lee, Melba M. Crawfork. Unsupervised multistage image classification using hierarchical clustering with a Bayesian similarity measure. *IEEE Trans. on Image Processing*, vol. 14, no. 3, Mar. 2005.

[31] Bjorn Stenger, Arasanathan Thayananthan, Philip H. S. Torr, Roberto Cippolla. Model-based hand tracking using a hierarchical Bayesian filter. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, no. 9, pp. 1372-1384, Sep. 2006.

Longcun Jin   received the B.S. degree in Telecommunication Engineering in 2005 from Hangzhou Dianzi University. He is currently a Ph.D. candidate in School of Communication and Information Engineering at Shanghai University. His research interests include in the field of data mining and virtual reality.

Wanggen Wan   received the Ph.D. degree in Electronic Engineering in 1992 from Xidian University. He is currently a professor in School of Communication and Information Engineering at Shanghai University. His research interests include virtual reality, data mining and machine learning. He is a Fellow of IET and a senior member of IEEE.