IEICE TRANS. INF. & SYST., VOL.E100–D, NO.2 FEBRUARY 2017

PAPER

Hierarchical Sparse Bayesian Learning with Beta Process Priors for Hyperspectral Imagery Restoration

Shuai LIU[†a)], ***Member***, **Licheng JIAO**[†], **Shuyuan YANG**[†], ***and*** **Hongying LIU**[†], ***Nonmembers***

SUMMARY Restoration is an important area in improving the visual quality, and lays the foundation for accurate object detection or terrain classification in image analysis. In this paper, we introduce Beta process priors into hierarchical sparse Bayesian learning for recovering underlying degraded hyperspectral images (HSI), including suppressing the various noises and inferring the missing data. The proposed method decomposes the HSI into the weighted summation of the dictionary elements, Gaussian noise term and sparse noise term. With these, the latent information and the noise characteristics of HSI can be well learned and represented. Solved by Gibbs sampler, the underlying dictionary and the noise can be efficiently predicted with no tuning of any parameters. The performance of the proposed method is compared with state-of-the-art ones and validated on two hyperspectral datasets, which are contaminated with the Gaussian noises, impulse noises, stripes and dead pixel lines, or with a large number of data missing uniformly at random. The visual and quantitative results demonstrate the superiority of the proposed method.

key words: *hierarchical sparse Bayesian learning, restoration, beta process, hyperspectral image*

1. Introduction

Hyperspectral images involve capturing of both spectral and spatial information by the modern sensors, organized band by band to constitute a three-dimensional (3D) data-cube. Due to the limitations of the acquisition systems, hyperspectral images unavoidably suffer from various degradations, such as noise contamination, stripe corruption, missing data [1], [2]. The degradations severely limit the quality of the HSI and influence the precision of the subsequent processing [3], including unmixing, target detection, classification and recognition. Hence, image restoration is of critical importance and challenging on the preprocessing of the HSI analysis.

In the literature, approaches based on wavelet transform are widely used for HSI restoration. They utilize some priors or regularizations to threshold the wavelet coefficients of the degraded data-cube in the transform domain, such as hybrid spatial-spectral derivatives [4], principal component analysis [5] or first order spectral roughness [6], and then perform the inverse wavelet transform to restore HSI. However, the proper kinds of wavelet transform are hard to choose. Considered the HSI as a three-order tensor, tensor analysis methods are proposed for noise reduction in [7] and [8], with the spectral and spatial information completely preserved.

Besides, nonlocal approaches can exploit spatial information on the whole image directly. By this, BM3D groups similar patches into 3D arrays and denoises these arrays by a sparse collaborative filtering [9]. In this way, the fine spatial structures of the image can be preserved. 3D nonlocal means (NLM) adoptes this idea and uses a nonlocal sparse model for restoration [10], [11]. Four-dimensional block matching (BM4D) is introduced to suppress the noises of the volumetric data in [12], with satisfying results.

Sparse and redundant representation over dictionary learning has been extensively introduced into HSI recovery with different types of regularization. K-means and K-singular value decomposition (KSVD) [13] adopts the local sparseness characteristic to remove additive Gaussian noise. Deep Dictionary Learning is explored in [14]. Due to the high correlation between hyperspectral signatures of pixels, the low rank constraint is incorporated as a regularization term to denoise the HSI [15].

Furthermore, the real HSI is usually contaminated by mixed noises, e.g., Gaussian noise, impulse noise, stripes, dead pixel lines, which are commonly not uniform across bands. Most techniques above only focus on identifying one or two kinds of the noises with specific prior knowledge, such as Gaussian noise with constant variance among the bands or impulse noise. Additionally, the hyperspectral images often face the loss of information during the acquiring process [16], [17], which is visible from acquired images as missing pixels. In order to ensure physical consistency in the restoration, the spectral characteristics of missing pixels need to be inferred, which is usually ignored in HSI restoration.

To mitigate the above limitations, we present a novel hierarchical framework to get the 'clean and entire' recovery images based on the sparse Bayesian learning, which develops the sparse representation based on dictionary learning [18], [19]. Beta process is employed as the independent sparse priors, and explored by use of Gibbs sampler. The posterior probability density function of variables can be efficiently sampled to infer the dictionary and the noise directly, with no prior knowledge of test data. It also makes prediction about both the mixed noises and the missing pixels in a data-driven manner simultaneously.

The paper is organized as follows. Section 2 gives a

Manuscript received August 1, 2016.
Manuscript revised September 28, 2016.
Manuscript publicized November 4, 2016.

[†]The authors are with Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, Joint International Research Laboratory of Intelligent Perception and Computation, Xidian University, Xi'an, Shaanxi Province, 710071, China.
a) E-mail: shliu_122908@yahoo.com

DOI: 10.1587/transinf.2016EDP7322

detailed description of the proposed restoration model, and describes the Bayesian inference which is performed with Gibbs sampler. Several experimental results are presented in Sect. 3 based on two HSI data. The conclusions are given in Sect. 4.

2. Hierarchical Sparse Bayesian Learning with Beta Process Prior

The sparse Bayesian learning (SBL) has been widely applied for the HSI analysis [20], [21], with successful results. But multi-dimensional integral needs to be calculated for solving sparse Bayesian problems, which is usually analytically intractable. Until now, the full conjugacy between the beta and Bernoulli distributions have been demonstrated [22]. So in the SBL model combining the beta and Bernoulli distributions, the posterior computation can be performed analytically. We start with a brief review of beta process in Sect. 2.1 and then provide a detailed description of the proposed model in Sect. 2.2.

2.1 Beta Process

The beta process is an infinite jump process, which is suitable for dictionary learning due to the high flexibility [23]. The two-parameter beta process is denoted by the draw $H \sim BP(a, b, H_0)$ with parameters a, $b > 0$, which is proposed in [24]. Let Ω be a measurable space and $\mathcal{B}$ its σ-algebra, the disjoint and infinitesimal partitions of Ω are denoted as $B \in \{B_1, \cdots, B_K\}$. The base measure H_0 is a fixed probability measure over $(\Omega, \mathcal{B})$ with $H_0(B_k) = \dfrac{1}{K}$ for $k = 1, \cdots, K$. The set function form of the two-parameter beta process is shown in Eq. (1).

$$\begin{aligned} H(B) &= \sum_{k=1}^{K} \boldsymbol{\pi}_k \delta_{B_k}(B) \\ \boldsymbol{\pi}_k &\sim Beta\left(\frac{a}{K}, \frac{b(K-1)}{K}\right) \end{aligned} \tag{1}$$

where H is composed by infinite number of B_k sampled i.i.d. from H_0 with K probabilities. $\boldsymbol{\pi}_k$ represents the jump, which is commonly utilized to parameterize a finite Bernoulli process. *Beta* defines the beta distribution. Supposing $\boldsymbol{z}_i \in R^K$ is drawn from a Bernoulli process with the parameter $\boldsymbol{\pi}_k$, $\boldsymbol{z}_i$ is a binary vector and can be written as $z_{ik} \sim Bernoulli(\boldsymbol{\pi}_k)$. In the dictionary learning, $\{B_k\}_{k=1,\cdots,K}$ refers to the dictionary atoms, and K represents the number of atoms. By reasonably choosing K, a and b, $\boldsymbol{\pi}_k$ will be near to zero and z_{ik} is equal to zero with a high probability, which implies the sparse constrains on the dictionary learning model.

2.2 Full Hierarchical Sparse Bayesian Model

Considering the HSI $Y \in R^{l_x \times l_y \times \lambda}$, where l_x and l_y define the size of the two spatial dimensions, and λ refers to the number of bands. To fully exploit highly correlated spectral information and strongly similar spatial information, Y is divided into overlapping 3D blocks instead of 2D blocks when performing the restoration. The size of each 3D block is $P = n_x \times n_y \times \lambda$, where $n_x \times n_y$ defines the spatial size of the 3D blocks. In vector form, each block is transformed into $\boldsymbol{x}_i \in R^P$ for $i = 1, \cdots, M$, and the total number of the 3D blocks is $M = (l_x - n_x + 1)(l_y - n_y + 1)$.

For HSI recovery, some existing noises, including impulse noises, stripes and dead pixel lines, only appear in small part of pixels within a band or few bands, and the intensity and positions of these noises are often more subtle and various. So these noises can be considered to be sparse in the hypersectral images. To fully depict the noise characteristics of HSI, we decompose the noise term into Gaussian noise term and sparse noise term.

$$\begin{aligned} &\boldsymbol{x}_i = \boldsymbol{D}\boldsymbol{\alpha}_i + \boldsymbol{n}_i + \boldsymbol{s}_i \circ \boldsymbol{v}_i \\ &\boldsymbol{d}_k \sim N(0, P^{-1}\boldsymbol{I}_P) \\ &\boldsymbol{\alpha}_i = \boldsymbol{z}_i \circ \boldsymbol{w}_i \\ &z_{ik} \sim Bernoulli(\boldsymbol{\pi}_k), \quad \boldsymbol{\pi}_k \sim Beta\left(\frac{a_\pi}{K}, \frac{b_\pi(K-1)}{K}\right) \\ &\boldsymbol{w}_i \sim N(0, \gamma_w^{-1}\boldsymbol{I}_K), \quad \gamma_w \sim \Gamma(c, d) \\ &\boldsymbol{n}_i \sim N(0, \gamma_n^{-1}\boldsymbol{I}_P), \quad \gamma_n \sim \Gamma(e, f) \\ &s_{ip} \sim Bernoulli(\theta_{ip}), \quad \theta_{ip} \sim Beta(a_\theta, b_\theta) \\ &v_{pi} \sim N(0, \gamma_v^{-1}), \quad \gamma_v \sim \Gamma(g, h) \end{aligned} \tag{2}$$

With these, the proposed model is considered as consisting of three terms. The first term represents the 'clean and entire' HSI, which can be well learned by elements of dictionary. The success attributes to the fact that valid data in corrupted images are intrinsically sparse under the dictionary framework and the noises are uniformly spread and cannot be represented by the dictionary. The second term is the Gaussian noise, and the third term defines the sparse noise. Beta process coupled with Bernoulli process is utilized to depict the sparseness of the valid data and the arbitrariness of the intensity and positions in the sparse noise. Gaussian process is exploited to learn the Gaussian noise. According to these, the 'clean and entire' image can be effectively restored from the degraded HSI while the noises can be greatly reduced by well learning their statistics characteristics. The proposed hierarchical restoration model is defined as Eq. (2).

In (2), the symbol $\circ$ represents the element-wise multiplication, $\boldsymbol{I}_P$ ($\boldsymbol{I}_K$) is the $P \times P$ ($K \times K$) identity matrix, and K is the number of the dictionary atoms. In this model, $D = [d_1, \cdots, d_k, \cdots, d_K] \in R^{P \times K}$ represents the dictionary learned from the test data, with the dictionary atoms drawn from a Gaussian distribution. The vector $\boldsymbol{\alpha}_i$ represents the sparse coefficient and $\mathbf{A} = [\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_i, \cdots, \boldsymbol{\alpha}_M]$ is the sparse coefficient matrix for $\{\boldsymbol{x}_i\}_{i=1,\cdots,M}$. The binary vector $\boldsymbol{z}_i = [z_{i1}, z_{i2}, \cdots, z_{iK}]^T$, drawn from a Bernoulli process, is coupled with π_k drawn from beta process. And it defines the columns of $\boldsymbol{D}$ exploited to represent $\boldsymbol{a}_i$ with probability π_k. The vector $\boldsymbol{w}_i = [w_{i1}, w_{i2}, \cdots, w_{iK}]^T$ is the weight of the coefficient $\boldsymbol{\alpha}_i$, which is learned by Gaussian process. When $K \to \infty$, the expectation of $\boldsymbol{z}_i$ is drawn from

Poisson(a_π/b_π) at random [24]. Therefore, explicit sparseness can be enforced on the coefficients $\{\boldsymbol{a}_i\}_{i=1,\cdots,M}$ through adjusting non-informative hyper-parameters a_π and b_π in beta process. When $z_{ik} = 0$, the coefficient a_{ik} is equal to zero instead of near zero in many sparse approaches, which means that the kth atom of $\boldsymbol{D}$ is not used for coefficient $\boldsymbol{\alpha}_i$. By calculating the number of the unused atoms, the size of dictionary can be inferred adaptively.

$N(\cdot)$ and $\Gamma(\cdot)$ represents the normal distribution and gamma distribution respectively, and these two distributions tender much more flexibility to solve the model with the posterior PDF. γ_w, γ_n and γ_v are taken for precision of parameters or noise precision separately, with a non-informative gamma prior.

In the sparse noise term, Beta-Bernoulli process and Gaussian distribution are explored to fully depict the arbitrariness of position and amplitude in the sparse noise separately. The intensity of sparse noise is defined by the matrix $\boldsymbol{v}_i = [v_{i1}, \cdots, v_{ip}, \cdots, v_{iP}]^T$, with each element constrained by the non-informative hyper-parameter priors. The matrix $\boldsymbol{s}_i = [s_{i1}, \cdots, s_{ip}, \cdots, s_{iP}]^T$ represents the location information of sparse noise in $\boldsymbol{x}_i$, which can be illustrated with non-zero elements of $\boldsymbol{s}_i$.

The negative logarithm posterior density function of the proposed method is represented as Eq. (3). According to (3), all observed and unknown variables can be considered as stochastic variables with the joint probability distribution specified. Therefore the proposed method is of more robustness and accuracy. In the proposed method, the distributions of all random variables are in the conjugate exponential family, and Gibbs sampler can be utilized to infer each variable by repeatedly sampling the conditional distributions. The detail inferences of Gibbs sampler are displayed in algorithm 1.

$$
\begin{aligned}
&-\log p(\{\boldsymbol{D}, \boldsymbol{W}, \boldsymbol{Z}, \boldsymbol{Q}, \boldsymbol{S}, \{\pi_k\}, \{\theta_{ip}\}, \gamma_w, \gamma_n, \gamma_v\}|\boldsymbol{X}) = \\
&0.5\gamma_n \sum\nolimits_i \|\boldsymbol{X} - \boldsymbol{D}(\boldsymbol{W} \circ \boldsymbol{Z}) - \boldsymbol{Q} \circ \boldsymbol{S}\|_2^2 + 0.5P \sum\nolimits_k \boldsymbol{d}_k^{\mathrm{T}} \boldsymbol{d}_k \\
&+ 0.5\gamma_w \sum\nolimits_i \|\boldsymbol{w}_i\|_2^2 + \sum\nolimits_k Beta\left(\pi_k \middle| \frac{a_\pi}{K}, \frac{b_\pi(K-1)}{K}\right) \\
&+ \sum\nolimits_{ik} \log Bernoulli(z_{ik}|\pi_k) \\
&+ 0.5\gamma_v \sum\nolimits_i \|\boldsymbol{q}_i\|_2^2 + \sum\nolimits_{ip} Beta(\theta_{ip}|a_\theta, b_\theta) \\
&+ \sum\nolimits_{ip} \log Bernoulli(s_{ip}|\theta_{ip}) \\
&+ \log \Gamma(\gamma_w|c, d)\Gamma(\gamma_n|e, f)\Gamma(\gamma_v|g, h) \\
&+ const
\end{aligned} \tag{3}
$$

Moreover, we observe $\boldsymbol{y}_i = \boldsymbol{\Sigma}_i \circ \boldsymbol{x}_i$ instead of $\boldsymbol{x}_i$ to make prediction for missing data by using the remain data, where $\boldsymbol{\Sigma}_i = \{0, 1\}^P$ is the sampling matrix with $\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_i^{\mathrm{T}} = \boldsymbol{I}_{\|\boldsymbol{\Sigma}_i\|_0}$. For $f = 1, \cdots, P$, $\boldsymbol{\Sigma}_{fi} = 0$ represents that the fth pixel of the vector $\boldsymbol{x}_i$ is lost.

Algorithm 1

Input: Noisy data $\boldsymbol{X}$, hyper-parameters
Output: Restored data $\boldsymbol{X}_{restore}$

Initialization: $Num = 100$, $K = 256$

for iter = 1 : Num
for $k = 1 : K$

Sampling $\boldsymbol{d_k}$: $p(\boldsymbol{d}_k|-) \sim N(\mu_{\boldsymbol{d}_k}, \boldsymbol{\Omega}_{\boldsymbol{d}_k})$

$$\boldsymbol{\Omega}_{\boldsymbol{d}_k} = \left(P\boldsymbol{I}_P + \gamma_n \sum_i w_{ik}^2 \boldsymbol{z}_{ik}^2\right)^{-1}$$

$$\boldsymbol{\mu}_{\boldsymbol{d}_k} = \gamma_n \boldsymbol{\Omega}_{\boldsymbol{d}_k} \sum_i w_{ik} \boldsymbol{z}_{ik} \boldsymbol{x}_{(i,-k)}$$

$$\boldsymbol{x}_{(i,-k)} = \boldsymbol{x}_i - \boldsymbol{D}(\boldsymbol{w}_i \odot \boldsymbol{z}_i) - \boldsymbol{q}_i \odot \boldsymbol{s}_i + (w_{ik} \odot \boldsymbol{z}_{ik})\boldsymbol{d}_k$$

Sampling z_{ik} : $p(\boldsymbol{z}_{ik}|-) \sim Bernoulli\left(\frac{p_1}{p_1 + p_0}\right)$

$$p_1 = \pi_k \exp(-0.5\gamma_n(w_{ik}^2 \boldsymbol{d}_k^T \boldsymbol{d}_k - 2w_{ik} \boldsymbol{d}_k^T \boldsymbol{x}_{(i,-k)}))$$
$$p_0 = 1 - \pi_k$$

Sampling $\boldsymbol{w_i}$: $p(w_{ik}|-) \sim N(w_{ik}|\mu_{wik}, \Omega_{wik})$

$$\Omega_{wik} = (\gamma_w + \gamma_n \boldsymbol{z}_{ik}^T \boldsymbol{d}_k^T \boldsymbol{d}_k)^{-1}$$
$$\mu_{wik} = \gamma_n \Omega_{wik} \boldsymbol{d}_k^T \boldsymbol{x}_{(i,-k)}$$

Sampling π_k:

$$p(\pi_k|-) \sim \mathrm{Beta}\left(\frac{a_\pi}{K} + \sum_i z_{ik}, \frac{b_\pi(K-1)}{K} + M - \sum_i \boldsymbol{z}_{ik}\right)$$

end
Sampling γ_w:
$p(\gamma_w|-) \sim \Gamma\left(c + 0.5MK, d + \sum_i 0.5 w_i^T w_i\right)$.
Sampling γ_n:
$p(\gamma_n|-) \sim \Gamma\left(e + 0.5PM, f + \sum\nolimits_i \|x_i - \boldsymbol{D}(\boldsymbol{w}_i \circ \boldsymbol{z}_i) - \boldsymbol{q}_i \circ \boldsymbol{s}_i\|_2^2\right)$
Sampling s_{ip}: $p(s_{ip}|-) \sim Bernoulli\left(\frac{v_1}{v_1 + v_0}\right)$
$v_1 = \theta_{ip} \exp(-0.5\gamma_v(q_{ip}^2 - 2q_{ip}\boldsymbol{x}_{(i,s)}))$, $v_0 = 1 - \theta_{ip}$
$\boldsymbol{x}_{(i,s)} = \boldsymbol{x}_i - \boldsymbol{D}(\boldsymbol{w}_i \circ \boldsymbol{z}_i)$
Sampling $\boldsymbol{q}_i$: $p(q_{ip}|-) \sim N(\mu_{qip}, \Omega_{qip})$
$\Omega_{qip} = (\gamma_v + \gamma_n s_{ip}^2)^{-1}$, $\mu_{qip} = \gamma_n \Omega_{qip} s_{ip} \boldsymbol{x}_{(i,s)}$
Sampling $\boldsymbol{\theta}_p$:
$p(\boldsymbol{\theta}_p|-) \sim \mathrm{Beta}\left(a_\theta + \sum_i s_{ip}, b_\theta + M - \sum_i s_{ip}\right)$
Sampling γ_v:
$p(\gamma_v|-) \sim \Gamma\left(g + 0.5PM, h + \sum_i 0.5 q_i^T q_i\right)$

end
Calculating $\boldsymbol{X}_{restore} = \boldsymbol{DA}$

Additionally, there are several solutions to the same spectral pixel due to the usage of overlapping 3D blocks, so the restored HSI is constructed by averaging all overlapping 3D blocks. Performing this operation for $\boldsymbol{X}_{restore}$, we can get the final HSI after restoration.

3. Experimental Results

The performance of the proposed restoration model is demonstrated on two hyperspectral datasets visually and quantitatively. One is Indian Pines data, with the spatial size of 145×145, which was acquired by Airborne Visible/Infrared Imaging Spectrometer in June 1992. It has 200 spectral bands of $10\,nm$ widths from $0.4\,\mu m$ to $2.45\,\mu m$ and a spatial resolution of $20\,m$. The second data set, Botswana, consists of 145 spectral wavelengths with 1476×256 pixels. It was acquired by the NASA EO-1 satellite with the Hyperion sensor on May 31, 2001. It is in $10\,nm$ windows with $30\,m$ spatial resolution over a $7.7\,km$ strip. A subset of size $150 \times 200 \times 145$ is used here. Note that noisy and water absorption bands were removed from both datasets in the experiment.

To investigate the performance of the proposed method, we choose four different methods for comparison, including K-SVD [6], BM3D [7], ANLM3D [11] and BM4D [12]. The necessary parameters in the four compared methods are finely tuned or automatically selected to generate the optimal simulated results. For the proposed method, the spatial size of blocks is 4×4, and spatial information can be employed by this in a certain degree. The size of dictionary and the iteration number of Gibbs sampling are set as $K = 128$ and 100 separately. The hyper-parameters of the Gaussian noise are set as $e = f = 10^{-6}$. For setting the remaining hyper-parameters, two cases are taken into consideration. 1) The HSI is contaminated by Gaussian noise, dead pixel lines or a mixture of them. The rest hyper-parameters are set as: $a_\pi = b_\pi = a_\theta = b_\theta = 10^{-6}$; $c = d = g = h = 10^{-6}$. 2) The HSI is contaminated by a mixture of Gaussian and impulse noise or a mixture of Gaussian noise, impulse noise and dead pixel lines. The rest hyper-parameters are set as: $a_\pi = b_\pi = a_\theta = b_\theta = 10^{-4}$; $c = d = g = h = 10^{-5}$. Once the types of noises are selected, the hyper-parameters will be also determined and have no necessary to be tuned.

The experimental results are evaluated in two ways. Firstly, visual comparisons are shown in the restored images and spectral signatures. Due to the huge amounts of data pixels and spectral bands, a few of them are presented in this paper. Secondly, peak signal to noise ratio (PSNR) is used to quantitatively measure the similarity between the restored and reference images based upon the mean square error. Structure similarity (SSIM) and feature similarity (FSIM) are utilized to measure structural consistency and perceptual consistency between each initial band and restored band respectively [25]–[27]. Normally, the higher the measure value is, the better quality the image has. The mean spectral angle (MSA) between different spectral pixels is employed to numerically evaluate spectral fidelity of the restored results [28], [29]. The MSA is calculated by Eq. (4).

$$MSA = \frac{1}{uv}\sum_{i=1}^{u}\sum_{j=1}^{v}\cos^{-1}\left(\frac{x_{ij}^{(c)T}\cdot x_{ij}^{(o)}}{\left\|x_{ij}^{(c)}\right\|\cdot\left\|x_{ij}^{(o)}\right\|}\right) \tag{4}$$

where $x_{ij}^{(c)}$ and $x_{ij}^{(o)}$ represent the restored spectral pixels and original spectral pixels located at (i, j), respectively. u and v are the number of pixels in the two spatial dimensions. T represents the transpose of matrix. Generally, the smaller the MSA values are, the better the spectral fidelity is. The following experiments consist of three subsections. Section 3.1 presents the restored results of the HSI polluted by various noises; Sect. 3.2 reports the inferring results of the HSI with some data missing, which also gives the results of the HSI degraded by noise contamination and missing data simultaneously; Sect. 3.3 explains the necessary of the sparse noise term.

3.1 Denoising

In the first simulated experiment, four kinds of noises are added to the datasets.

1) Zero-mean Gaussian noise is added to all bands, with the noise variance σ fixed or varying across bands randomly.

Tables 1 and 2 show the PSNR values of restored results with five different approaches and Tables 3 and 4 display the MSA values before and after denoising, in which the two simulated datasets are corrupted by Gaussian noise with the noise standard deviation $\sigma = [5, 15, 25, 35, 50]$. The best measure values are bolded to improve the comparisons. Clearly, the proposed method has the highest PSNR values and the lowest MSA values than other four compared

Table 1 PSNR comparison with different methods for Indian Pines data

	PSNR				
	5	15	25	35	50
Corrupted HSI	34.1463	24.6095	20.174	17.7596	14.1553
KSVD	35.8408	31.24	27.8295	25.7287	22.3953
BM3D	38.6757	33.1185	30.8745	29.6837	28.498
ANLM3D	38.8343	33.9009	31.5687	29.962	28.7341
BM4D	41.18	35.5003	33.8228	32.1876	30.5854
Ours	**42.8863**	**39.3473**	**37.3712**	**34.5684**	**32.4548**

Table 2 PSNR comparison with different methods for Botswana data

	PSNR				
	5	15	25	35	50
Corrupted HSI	34.1473	24.6107	20.1724	17.252	14.1511
KSVD	35.9896	30.8125	27.6338	25.1636	25.6029
BM3D	37.6772	32.3446	30.1449	29.0141	27.0309
ANLM3D	39.6791	33.9057	31.6826	30.0964	28.902
BM4D	42.1063	36.2804	33.6897	32.0322	30.7741
Ours	**43.3222**	**39.2766**	**37.0629**	**35.5206**	**33.6706**

Table 3 MSA comparison with different methods for Indian Pines data

	MSA				
	5	15	25	35	50
Corrupted HSI	0.0277	0.0829	0.1378	0.1917	0.2704
KSVD	0.0279	0.0341	0.0434	0.0541	0.0718
BM3D	0.0243	0.0303	0.0362	0.0428	0.0579
ANLM3D	0.0193	0.0297	0.038	0.0454	0.0558
BM4D	0.0138	0.0259	0.0299	0.0352	0.0416
Ours	**0.0140**	**0.0211**	**0.0249**	**0.0293**	**0.0357**

Table 4 MSA comparison with different methods for Botswana data

	MSA				
	5	15	25	35	50
Corrupted HSI	0.0755	0.2208	0.3526	0.46907	0.6161
KSVD	0.0445	0.0914	0.1435	0.1957	0.1336
BM3D	0.0447	0.0813	0.0886	0.1073	0.147
ANLM3D	0.0365	0.0605	0.0797	0.0981	0.1267
BM4D	0.0299	0.0549	0.0626	0.0705	0.0836
Ours	**0.0227**	**0.0368**	**0.0441**	**0.0504**	**0.0599**

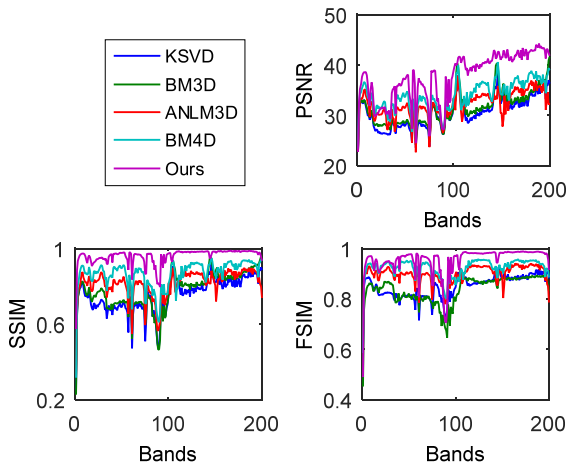

Fig. 1 PSNR, SSIM and FSIM values of each band with different methods for Indian Pines data

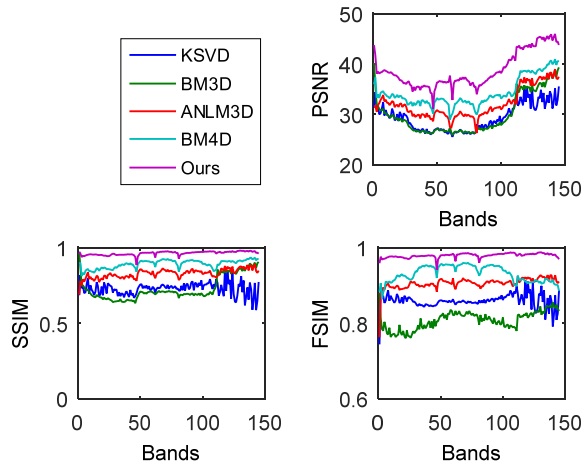

Fig. 2 PSNR, SSIM and FSIM values of each band with different methods for Botswana data

ones. This means that our method can better promote the image quality and preserve the spectral information on each located pixel while removing Gaussian noise. According to Tables 1–4, KSVD and BM3D generate the worse values than other methods, which are attributed to the fact that KSVD and BM3D denoise the HSI band by band and destroy the spectral correlations. This inferiority is much obvious in estimating spectral fidelity as shown in Tables 3 and 4. ANLM3D has a higher PSNR values than KSVD and BM3D, but it does worse work than BM3D in some MSA values. BM4D has the better values than KSVD, BM3D and ANLM3D, but it is obviously inferior to the proposed method.

Figures 1 and 2 display the performance curves of PSNR, SSIM and FSIM values for India Pines data and Botswana data separately, in which noise standard deviations change across bands within the interval $[15, 30]$. In Figs. 1 and 2, the curves have the obvious fluctuations due to the varying σ with the bands. It is easily found that the performances of different algorithms are almost the same at some bands, which arises from the fact that much smaller Gaussian noise has been added to these bands than their adjacent bands.

Obviously, the PSNR, SSIM and FSIM values with the proposed approach are higher than the competitors at most bands, which have a more stable trend at the same time. This is because the proposed method can well learn noise characteristics and adaptively infer the noise standard deviation. KSVD and BM3D show the lower values in both Fig. 1 and Fig. 2. And they denoise all bands with fixed noise levels, which fail to dispose the structural and feature information on degraded images. Exploring both spatial and spectral information, ANLM3D and BM4D yield the higher measure results than KSVD and BM3D. And both of them present the unstable performance as shown in Fig. 1 and Fig. 2. By sampling the infinite parameter space, the proposed method can obtain the optimum solution to the HSI recovery whether the noise standard deviation within each band is equal or not. For one Gibbs sampling iteration, the computational complexity of the proposed method is near to $O(K(P+M)+PM)$. And it should be pointed out that the suggested method consumes more time than the four compared ones.

2) Impulse noise with the deviation from 0.01 to 0.02 is added to 10 bands selected randomly.

3) Dead pixel lines are simulated for the randomly selected bands. The width of the dead pixel lines is from one line to three lines. In the following experiment, we add dead pixel lines to eight bands in the same position separately, which are from band 43 to band 46 and from band 129 to band 132.

4) Stripes are added to the randomly selected bands. The width of the stripes is from one line to three lines. Due to the similarity between dead pixel lines and stripes, we leave out the presentation of results for stripes in this work.

Figure 3 presents the images of band 45 for Indian Pines data after the removal of the mixed impulse noise and dead pixel lines. Also, we consider that hyperspectral images are polluted by the mixed Gaussian noise, impulse noise and dead pixel lines. The restoration results of band 130 for Indian Pines data are shown in Fig. 4.

It can be easily observed that the proposed method achieves outstanding performance in the visual results. KSVD employs the iterative method to learn the dictionary adaptively and improves the image quality greatly compared with the corrupted ones. But KSVD learns the dictionary atoms one by one and destroys the structures of sparse coefficients, which results in the loss of the edges and other structural details as shown in Fig. 3 (c) and Fig. 4 (c).

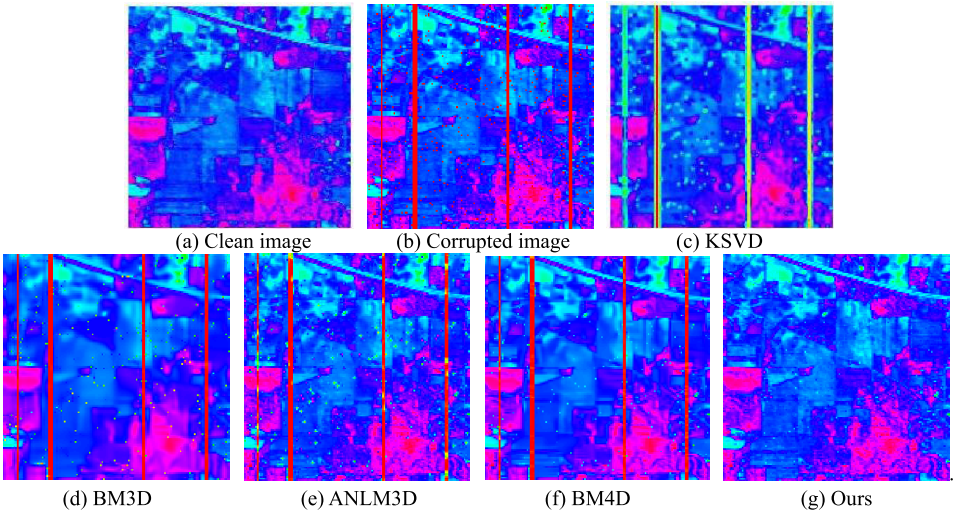

Fig. 3 Restoration of band 45 with a mixture of impulse noise and dead pixel lines

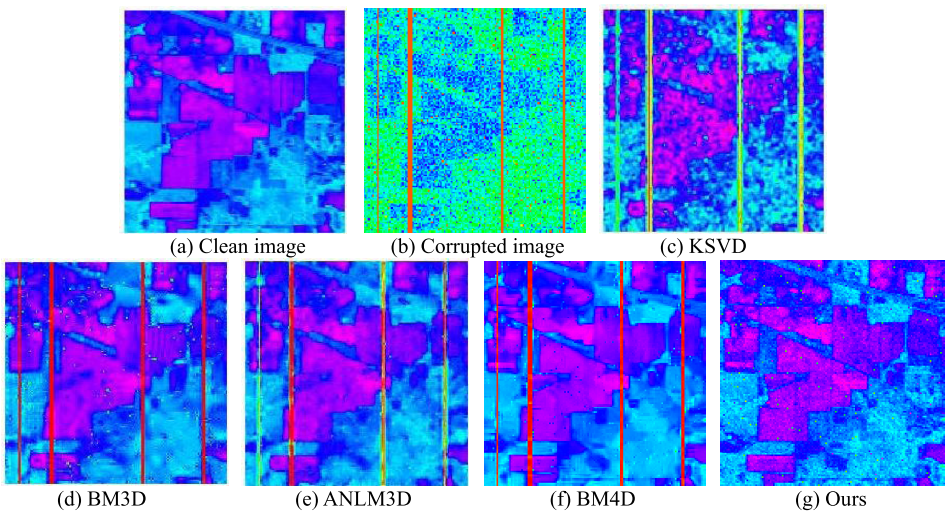

Fig. 4 Restoration of band 130 with the mixed Gaussian noise, impulse noise and dead pixel lines

Meanwhile, it needs carefully parameter tuning, e.g., noise level, the size of dictionary and blocks. Due to the highly spatial consistencies (see in Fig. 3 (a) and Fig. 4 (a)), BM3D has a large amount of similar blocks to achieve the restoration. By computing the average of the similar noisy blocks, it can effectively smooth the noises and makes a much better visual impression than KSVD. However, BM3D smoothes out some image structures and details while performing the recovery, and it is highly sensitive to noise level. Both K-SVD and BM3D are the bandwise approaches, which neglect the spectral continuity and correlations in HSI.

ANL3D can effectively utilize the high nonlocal self-similarity to better balance smoothing and details preservation. It fails to preserve the edges and some of the fine details for recovering seriously degraded images, in which local image structure be corrupted heavily. BM4D can achieve visual improvements by adopting the three-dimensional nonlocal self-similarity data cube. Using the BM4D method, the high spectral correlations between the continuous bands are not fully exploited; instead, only local correlations between some neighboring bands are explored. And its results smooth out some fine details.

There are obvious dead pixel lines as displayed in Fig. 3 (c)–(f) and Fig. 4 (c)–(f), which means the four compared approaches fail in restoring the seriously degraded HSI. According to Fig. 3 (g) and Fig. 4 (g), the obvious superiority of the proposed method can be easily found in detail preservation and the mixed noise reduction, which imply that the proposed approach is of possibility to predict the miss pixels. Generally, the proposed approach achieves more promising denoising performance, which is in line with the quantitative results in Tables 1–4, Fig. 1 and Fig. 2.

3.2 Predicting the Missing Data

In this subsection, 2% of the test data are randomly observed

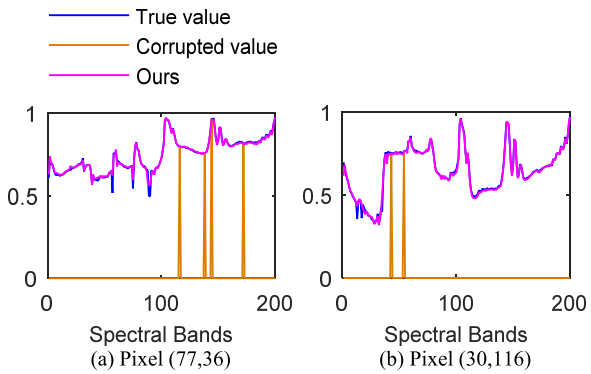

Fig. 5 Spectrum of different pixels for Indian Pines data

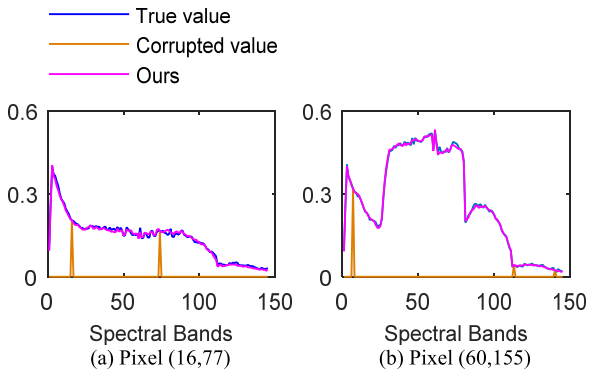

Fig. 6 Spectrum of different pixels for Botswana data

to estimate the performance of the proposed method, and in other words, 98% of the test data are missing. Then the full HSI is recovered by employing the 2% data. Figure 5 shows the true spectral signatures, the corrupted spectral signatures and inferred spectral signatures at different pixels, based on randomly observing 2% of Indian Pines data. In Fig. 6, the same is done for Botswana data. Noticing that, in the curves of corrupted values, the values of missing data are equal to zero, while the values of observed data are larger than zero. It is visually clear that the inferred spectra are very near to the true values. This means that the proposed method can efficiently restore the spectrum of HSI with very little data observed, and offers the capability of identifying and terrain classification for HSI analysis.

Furthermore, we consider the more realistic case that the HSI is degraded by noise contamination and missing data simultaneously. And Botswana data is utilized for simulating in this case, which is degraded by a mixture of Gaussian noise with standard deviation of 25 and the missing of 98% data. From the aspect of visual impressions, Fig. 7 displays the restored results, where the false-color images are represented by band combination of 85 (red), 36 (green) and 70 (blue). According to Fig. 7, the suggested method shows convincible results with greatly preserving the structure and detail information. Above all, the proposed method has the great superiority in both the spectral signatures preservation and the visual appearance for predicting the missing pixels.

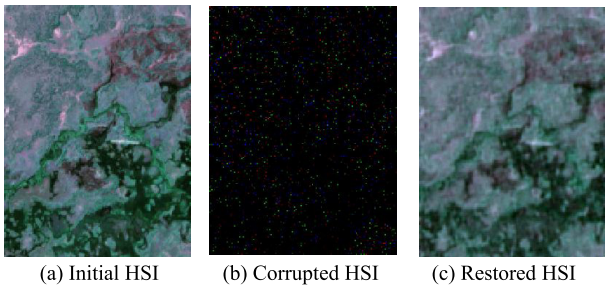

Fig. 7 Recovery image of Botswana data with Gaussian noise standard deviation of 25 and 98% data missing

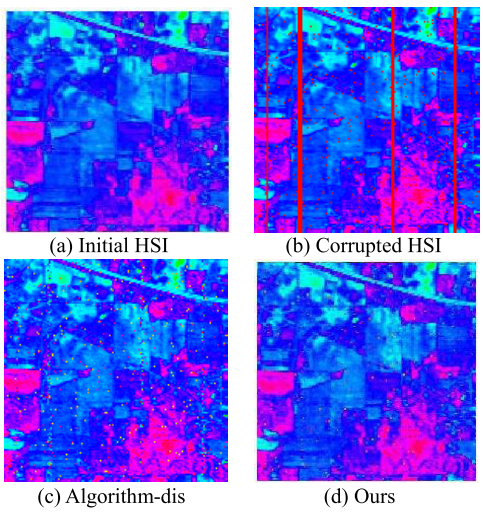

Fig. 8 Restoration of band 45 with a mixture of impulse noise and dead pixel lines

3.3 Discussion

To explain the necessary of sparse noise terms, we consider the Indian Pines data in this subsection, which is polluted by the impulse noise with the deviation 0.02 and dead pixel lines. The method generated by disabling sparse noise term of the proposed one is named as Algorithm-dis. As presented below, Fig. 8 shows the visual impression of band 45 obtained by the proposed method and Algorithm-dis. Figure 9 shows the horizontal profiles of band 45 at location $(115, 30)$, and Fig. 10 displays the spectral signatures at location $(115, 30)$.

According to Fig. 8, it can be easily found that both of the algorithms improve the image quality greatly compared with the corrupted ones in Fig. 8 (b). The Algorithm-dis can only reduce part of the impulse noise, as presented in Fig. 8 (c). And it fails in preserving the fine objects. As presented in Fig. 8 (d), the proposed method can effectively remove the impulse noise and dead pixel lines while preserving the local details such as edges and textures. Obviously, the proposed method achieves the better performance than Algorithm-dis.

After the denoising processing, the fluctuations are re-

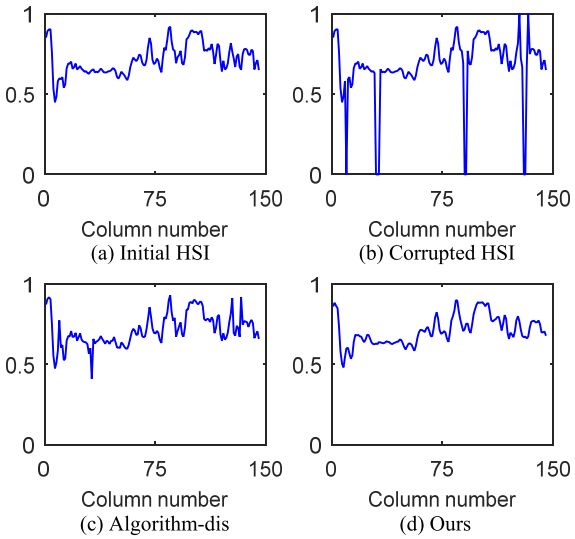

Fig. 9 The horizontal profiles of band 30 at location (115, 30)

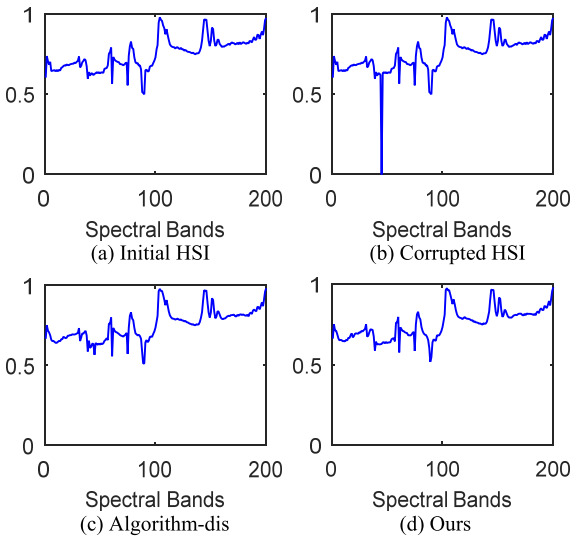

Fig. 10 The spectral signatures at location (115, 30)

duced in some degree. As shown in Figs. 9–10, the results obtained by the proposed method are closest to the curves of true value, which means the proposed method can greatly remove the mixed noise while well preserving the edge and texture information. From Figs. 9–10 (b), it can be observed that the curves have some obvious fluctuations, due to the existence of impulse noise and dead pixel lines. The results calculated by Algorithm-dis fail in effectively restoring the shape and amplitude of the clean HSI and loses some details, which can be seen from Figs. 9–10 (c).

4. Conclusion

In this paper, a hierarchical sparse Bayesian learning method is developed to address the HSI restoration problem by efficiently learning and representing the valid data and the noise characteristics. The conjugacy of the model posterior is explored by using the beta process, which is exploited to adaptively manifest the sparseness. The Gibbs sampler is then applied to infer the HSI restoration model. Unlike most studies in the literature, which considered Gaussian noises alone in the analysis, we learned the mixed noise while adopting the spatial information.

By comparison with the state-of-the-art methods, the experimental results illustrate the effectiveness of the proposed method on simultaneously and greatly reducing the mixture of Gaussian noise, impulse noise, dead pixel lines or stripes. And it also demonstrates its accuracy in predicting the missing data. Further research will be directed towards learning the hidden structure of the test data as the priors of the model.

Acknowledgments

This work is supported by the National Basic Research Program of China (973Program) with no. 2013CB329402, the Program for Cheung Kong Scholars and Innovative Research Team in University with No. IRT_15R53 and the Major Research Plan of the National Natural Science Foundation of China (Grant No. 91438103).

References

[1] B. Aiazzi, L. Alparone, A. Barducci, S. Baronti, P. Marcoionni, I. Pippi, and M. Selva, "Noise modelling and estimation of hyperspectral data from airborne imaging spectrometers," Annals of Geophysics, vol.49, no.3, pp.1–9, Feb. 2006.

[2] A.N. Zemliachenko, R.A. Kozhemiakin, M.L. Uss, S.K. Abramov, N.N. Ponomarenko, V.V. Lukin, B. Vozel, and K. Chehdi, "Lossy compression of hyperspectral images based on noise parameters estimation and variance stabilizing transform," J. Appl. Remote Sens, vol.8, no.1, pp.083571–083571, Aug. 2014.

[3] H. Peng, Automatic Denoising and Unmixing in Hyperspectral Image Processing, Thesis, Rochester Institute of Technology, 2014.

[4] H. Othman and S.-E. Qian, "Noise reduction of hyperspectral imagery using hybrid spatial-spectral derivative-domain wavelet shrinkage," IEEE Trans. Geosci. Remote Sens., vol.44, no.2, pp.397–408, Feb. 2006.

[5] G. Chen and S.-E. Qian, "Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage," IEEE Trans. Geosci. Remote Sens., vol.49, no.3, pp.973–980, March 2011.

[6] B. Rasti, J.R. Sveinsson, M.O. Ulfarsson, and J.A. Benediktsson, "Hyperspectral image denoising using first order spectral roughness penalty in wavelet domain," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol.7, no.6, pp.2458–2467, June 2014.

[7] X. Liu, S. Bourennane, and C. Fossati, "Reduction of signal-dependent noise from hyperspectral images for target detection," IEEE Trans. Geosci. Remote Sens., vol.52, no.9, pp.5396–5411, Sept. 2014.

[8] X. Guo, X. Huang, L. Zhang, and L. Zhang, "Hyperspectral image noise reduction based on rank-1 tensor decomposition," ISPRS J. Photogramm Remote Sens., vol.83, pp.50–63, Sept. 2013.

[9] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," IEEE Trans. Image Process., vol.16, no.8, pp.2080–2095, Aug. 2007.

[10] Y. Qian and M. Ye, "Hyperspectral imagery restoration using nonlocal spatial-spectral structured sparse representation with noise estimation," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol.6, no.2, pp.499–515, April 2013.

[11] J.V. Manjón, C. Pierrick, M.B. Luis, D.L. Collins, and M.

Robles, "Adaptive non-local means denoising of MR images with spatially varying noise levels," J. Magn. Reson. Imaging, vol.31, no.1, pp.192–203, Jan. 2010.
[12] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "Nonlocal transform-domain filter for volumetric data denoising and reconstruction," IEEE Trans. Image Process., vol.22, no.1, pp.119–133, Jan. 2013.
[13] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," IEEE Trans. Image Process., vol.54, no.11, pp.4311–4322, Nov. 2006.
[14] L.G. Huo, X.C. Feng, C.L. Huo, and C.H. Pan, "Learning Deep Dictionary for Hyperspectral Image Denoising," IEICE Trans. Inf. & Syst., vol.E98-D, no.7, pp.1401–1404, July 2015.
[15] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," IEEE Trans. Geosci. Remote Sens., vol.52, no.8, pp.4729–4743, Aug. 2014.
[16] K. Degraux, V. Cambareri, L. Jacques, B. Geelen, C. Blanch, and G. Lafruit, "Generalized inpainting method for hyperspectral image acquisition," IEEE International Conference on Image Processing, Quebec, Canada, pp.315–319, Sept. 2015.
[17] D. Cerra, R. Muller, and P. Reinartz, "Unmixing-based denoising for destriping and inpainting of hyperspectral images," 2014 IEEE Geoscience and Remote Sensing Symposium, Quebec, Canada, pp.4620–4623, July 2014.
[18] A. Shah, D.A. Knowles, and Z. Ghahramani, "An empirical study of stochastic variational algorithms for the beta bernoulli process," Proc. 32nd International Conf. on ICML, Lille, France, July 2015.
[19] D. Dranishnikov, Bayesian hyperspectral unmixing with multivariate beta distributions, Universityof FLORIDA, 2014.
[20] L. Sun, Z. Wu, J. Liu, L. Xiao, and Z. Wei, "Supervised spectral–spatial hyperspectral image classification with weighted Markov random fields," IEEE Trans. Geosci. Remote Sens., vol.53, no.3, pp.1490–1503, March 2015.
[21] L. Xu, F. Li, A. Wong, and D.A. Clausi, "Hyperspectral Image Denoising Using a Spatial–Spectral Monte Carlo Sampling Approach," IEEE J. Sel. Topics Appl. Earth Observ., vol.8, no.6, pp.3025–3038, June 2015.
[22] R. Thibaux and M.I. Jordan, "Hierarchical beta processes and the Indian buffet process," Proc. International Conf. on AISTATS, San Juan, Puerto Rico, vol.2, pp.564–571, March 2007.
[23] L. He, H. Qi, and R. Zaretzki, "Beta process joint dictionary learning for coupled feature spaces with application to single image super-resolution," Proc. International Conf. on CVPR, Portland, United States of America, pp.345–352, June 2013.
[24] J. Paisley and L. Carin, "Nonparametric factor analysis with beta process priors," Proc. 26th International Conf. on ICML, Montreal, Canada, June 2009.
[25] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. Image Process., vol.13, no.4, pp.600–612, April 2004.
[26] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A Feature Similarity Index for Image Quality Assessment," IEEE Trans. Image Process., vol.20, no.8, pp.2378–2386, Aug. 2011.
[27] F. Deger, A. Mansouri, M. Pedersen, et al., "A sensor-data-based denoising framework for hyperspectral images," Opt. Lett., vol.23, no.3, pp.1938–1950, March 2015.
[28] Q. Yuan, L. Zhang, and H. Shen, "Hyperspectral image denoising with a spatial–spectral view fusion strategy," IEEE Trans. Geosci. Remote Sens., vol.52, no.5, pp.2314–2325, May 2014.
[29] J. Chen, X. Jia, W. Yang, and B. Matsushita, "Generalization of subpixel analysis for hyperspectral data with flexibility in spectral similarity measures," IEEE Trans. Geosci. Remote Sens., vol.47, no.7, pp.2165–2171, July 2009.

Shuai Liu received the B.S. degree in Observation & Control Engineering and Instrumentation from Xidian University, Xi'an, China, in 2005. She is currently working toward the Ph.D. degree in the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, School of Electronic Engineering, Xidian University, Xi'an, China. Her major research interests include statistical machine learning and deep learning.

Licheng Jiao received the B.S. degree from Shanghai Jiaotong University, Shanghai, China, in 1982 and the M.S. and Ph.D. degree from Xian Jiaotong University, Xian, China, in 1984 and 1990. He is currently a Distinguished Professor at the School of Electronic Engineering, Xidian University, Xi'an. He has led approximately 40 important scientific research projects and has published more than 10 monographs and 100 papers in international journals and conferences. He is the author or coauthor of more than 150 scientific papers. Prof. Jiao is a member of the IEEE Xian Section Executive Committee and the Chairman of the Awards and Recognition Committee and an executive committee member of the Chinese Association of Artificial Intelligence. His research interests include signal and image processing, natural computation, and intelligent information processing.

Shuyuan Yang received the B.A. degree in electrical engineering from Xidian University, Xi'an, China, in 2000, the M.S. degree and Ph.D. degree in Circuit and System from Xidian University, Xi'an, China, in 2003 and 2005, respectively. She is now a professor at the School of electrical engineering in Xidian University and her main current research interests are machine learning and multiscale geometric analysis.

Hongying Liu received her B.E. and M.S. degrees in Computer Science and Technology from Xi'An University of Technology, China, in 2006 and 2009, respectively, and Ph.D. in Engineering from Waseda University, Japan in 2012. Currently, she is a faculty member at the School of Electronic Engineering, and also with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, Xidian University, China. In addition, she is a member of IEEE. Her major research interests include image processing, machine learning, compressive sampling, etc.