Fusion of Hyperspectral and Multispectral Images Based on a Bayesian Nonparametric Approach

Lichun Sui, Li Li, Jonathan Li, *Senior Member, IEEE*, Nan Chen, and Yongqing Jiao

***Abstract*—This paper presents a new approach to fusion of hyperspectral and multispectral images based on Bayesian nonparametric sparse representation. The approach formulates the image fusion problem within a constrained optimization framework, while assuming that the target image lives in a lower dimensional subspace. The subspace transform matrix is determined by principal component analysis, and the sparse regularization term is designed depending on a set of dictionaries and sparse coefficients associated with the observed images. Specifically, the dictionary elements and sparse coefficients are learned by the Bayesian nonparametric approach with the beta-Bernoulli process, which establishes the probability distribution models for each latent variable and calculates the posterior distributions by Gibbs sampling. Finally, serving the obtained posterior distributions as *a priori*, the fusion problem is solved via an alternate optimization process, where the alternate direction method of multipliers is applied to perform the optimization with respect to the target image. The Bayesian nonparametric method is used to optimize the sparse coefficients. Exhaustive experiments using both two public datasets and one real-world dataset of remote sensing images show that the proposed approach outperforms the existing state-of-the-art methods.**

***Index Terms*—Bayesian nonparametric model, dictionary learning, hyperspectral and multispectral images, image fusion, sparse representation.**

I. Introduction

In RECENT years, hyperspectral images (HSIs) have been widely used in a range of applications, such as monitoring and management of natural resources, biodiversity, ecosystems, and disasters [1]–[4]. HSIs, which have the characteristics of high spectral resolution, acquire a faithful spectral representation of a scene. However, because of the physical and technical bottlenecks of imaging spectrometers, HSIs are relatively restricted in terms of signal-to-noise ratio (SNR), spatial resolution, and scanning width, which, to some extent, limit the range of potential applications, such as medical imaging and remote sensing [5].

If high spatial resolution images, such as multispectral images (MSIs) of a scene of interest are available, an HSI with high spatial resolution is obtained by using an image fusion approach, which effectively combines the multiband information of a single sensor, or the single-band information of multiple sensors, thereby improving the precision and effectiveness of interpretation. Fusion data is viewed as a product of a synthetic sensor consisting of the high spectral resolution of the hyperspectral sensor and the high spatial resolution of the multispectral sensor, which allows for additional new potential applications, such as high spatial resolution ecosystem monitoring, or the monitoring of urban surface materials, minerals, etc. [6].

Considering remote sensing images, pan-sharpening (archetypal fusion) generally refers to the fusion of lower spatial resolution MSIs and a higher spatial resolution panchromatic (PAN) image. Pan-sharpening has been explored for several decades. The representative techniques of pan-sharpening are classified roughly as follows: component substitution (CS) [7], [8], multiresolution analysis (MRA) [9], [10], and sparse representation (SR) [11]–[13]. Because pan-sharpening can be regarded as a special case of HSI–MSI fusion, much effort has been made toward generalizing the existing pan-sharpening technique for HSI–MSI fusion. A pan-sharpening approach based on a generalized Laplacian pyramid (GLP) was used in [10] for HSI–MSI fusion. Although the algorithm could not de-noise, its performance was associated with noise post-processing. A hybrid pan-sharpening hyperspectral method was used in [14] for fusing a pan-sharpening image with an HSI.

Indeed, because HSI–MSI fusion contains more spatial and spectral information, it differs from pan-sharpening methods, thereby rendering many pan-sharpening methods, such as component replacement [7], [8] and relative spectral contribution [15] inapplicable or inefficient for HSI–MSI fusion. In this regard, more sophisticated attempts with respect to HSI–MSI/PAN fusion have been explored. Based on recent advances in pan-sharpening, a framework called hypersharpening was developed through adapting the MRA-based pan-sharpening methods to HSI–MSI fusion [16]–[18]. The main idea of the framework is to synthesize a high-resolution image for each

Manuscript received October 10, 2018; revised January 28, 2019; accepted February 21, 2019. This work was supported in part by the National Natural Scientific Foundation of China under Grants 41372330, 41601345, 41571346, 41871380, and 4141379. *(Corresponding author: Li Li.)*

L. Sui and N. Chen are with the College of Geology Engineering and Geomatics, Changan University, Xian 710054, China (e-mail: sui1011@chd.edu.cn; chcdut@126.com).

L. Li is with the JIKAN Research Institute of Engineering Investigations and Design, Co., Ltd., Xi'an 710043, China, and also with the College of Geology Engineering and Geomatics, Changan University, Xian 710054, China (e-mail: 2014026013@chd.edu.cn).

J. Li is with the Departments of Geography and Environmental Management, and Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada, and also with the Fujian Key Laboratory of Sensing and Computing for Smart Cities, School of Informatics, Xiamen University, Xiamen 361005, China (e-mail: junli@uwaterloo.ca).

Y. Jiao is with the Department of Surveying and Mapping, Gansu Industry Polytechnic College, Tianshui 741025, China (e-mail: 527644909@qq.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2019.2902847

HSI band as a linear combination of MSI bands via linear regression [16]. In addition, the principle of spectral unmixing was applied to multisensor, multiresolution image fusion [19], [20]. In [19], an approach was presented to unmix HSIs with low spatial resolution and enhance abundance maps by merging HSIs with higher resolution PAN images using constrained optimization algorithms. A method was presented in [20] to obtain the spatial details of segmentation from high-resolution images, then use them to unmix the low-resolution images, and finally sharpen the low-resolution images by assigning estimated endmember signatures to the related high-resolution pixels of the segmentation map. A nonnegative matrix factorization (NMF) based HSI–MSI fusion algorithm was presented in [21] for the midinfrared bands. The HSI with low spatial resolution is unmixed by NMF. Using least squares regression by holding the resampled endmember signatures fixed, the abundance maps with high spatial resolution are acquired from the MSI. A coupled nonnegative matrix factorization (CNMF) unmixing was used in [22] for the HSI–MSI fusion based on NMF, where both HSIs and MSIs were alternately unmixed into abundance and endmember matrices by the CNMF approach. Later, a similar fusion algorithm based on coupled spectral unmixing (CSU) was presented in [23]. A coupled sparse nonnegative matrix factorization method was constructed in [24] for the fusion of PAN and MSIs. To solve the fusion and unmixing problem by maximizing the joint posterior distribution of endmember and abundance matrices, a multiband image fusion method was presented in [25] based on spectral unmixing. Recently, two joint-criterion nonnegative matrix factorization approaches were proposed in [18] for hypersharpening, yielding sharpened HSIs with good spectral and spatial fidelities. A hybrid unmixing algorithm was proposed in [26] for HSIs based on region adaptive segmentation. A multiple multiband images fusion (MMIF) method was presented in [27] using the well-known forward observation and linear mixture models to cast the fusion problem as a reduced-dimension liner inverse problem, which is more suitable for the fusion of more than two multiband images.

Recently, due to its ability to exploit the self-similarity properties of natural images, sparse representation has received much attention [28]–[31]. Based on this property, sparse constraints were applied in both [32] and [33] to regularize the ill-posed super-resolution and/or fusion problems. An HSI restoration method was created in [34] to measure the low-rank structure with a new sparsity regularizer. In [33], multiple images were first decomposed into high- and low-frequency components followed by fusing them through applying a sparse regularization. However, this method requires a training dataset to learn the dictionary in advance. An online coupled dictionary learning based pan-sharpening method was presented in [35]. The Bayesian SR (BSR) algorithm was used in [36] to learn the endmember and abundance matrices, which, by super-resolution, implement the HSI–MSI fusion. An approach based on dictionary learning and sparse coding was presented in [37] to obtain the endmember and abundance matrices, respectively. However, the performance of the method proposed in [37] was sensitive to the algorithm parameters, especially to the dimensions of matrices, such as dictionary size. There is no principled method to introduce prior knowledge to enhance the performance. Later, a Bayesian-based HSI–MSI fusion method was presented in [38]. The problem formulation is based on information of the prior distribution in the observed scene, such as Gaussian or sparsity promoted Gaussian. Both subspace transformation and regularization are used in the fusion problem to solve the ill-conditioned inverse problem. On this basis, a fast fusion method was developed in [39] based on Sylvester equation (FUSE) method through integrating a Sylvester equation-based explicit solution into the Bayesian HSI–MSI fusion method. The FUSE method significantly reduces the computational complexity, while obtaining the same performance as the approach proposed in [39]. An HSI–MSI fusion algorithm was presented in [40] based on sparse representation (SRFM). Their method, to learn the dictionary and sparse coefficients of the observed images projected into the low-dimensional subspace, used Online Dictionary Learning and Orthogonal Matching Pursuit algorithms, respectively. Conditionally, with these dictionaries and sparse coefficients, the fusion problem was solved via alternate optimization with respect to the target image and the sparse coefficients. This approach shows the state-of-the-art results in HSI–MSI fusion. Although very powerful, the algorithm presented in [40] required one to specify the number of the dictionary elements and the variance of the noise in advance parameters that may be difficult to assess for real-world images. It was also unable to introduce more priori information during the dictionary learning and sparse coefficients learning and update to further improve the quality of the fusion results. Bayesian nonparametric algorithms can circumvent all the limitations. These algorithms adapt sufficiently the structure of the latent space to the images and infer the algorithm parameters nonparametrically that otherwise have to be assigned *a priori* [41]. The nonparametric properties decrease the human interference factors and introduce uncertainty by applying probabilistic estimation instead of point estimation with respect to the unknown parameters in the estimation process, further improving the automation degree of the models and enhancing the performance of fusion methods [42]. In this regard, we develop an HSI–MSI fusion method based on Bayesian nonparametric sparse representation in this paper.

The proposed approach fuses HSIs with MSIs within a constrained optimization framework roughly in a two-stage process.

Stage 1: In the first stage, the proposed approach projects the observations, i.e., HSIs and MSIs, into a low-dimensional subspace, and then regards them as a training dataset to learn the overcomplete dictionary and sparse coefficients using the Bayesian nonparametric method with beta process [41]. More specifically, the beta-Bernoulli process is applied to establish the probability distribution models of the dictionary elements and other parameters, while the Gibbs sampling method [43] is used to calculate the posterior distribution of each latent variable.

Stage 2: In the second stage, the proposed approach solves the final constrained optimization problem by alternatively optimizing with respect to the target image and the corresponding sparse coefficients. The Split Augmented Lagrangian Shrinkage Algorithm (SALSA) [44], an instance of the Alternating Direction Method of Multipliers (ADMM), is performed to achieve

optimization with respect to the target image. The Bayesian nonparametric approach is used to optimize the sparse coefficients, where the posterior distributions of the latent variables obtained from the first stage are served as *a priori* information in this stage.

The rest of this paper is structured as follows. Section II formulates the image fusion problem in a constrained optimization framework. Section III presents the proposed method of dictionary, sparse coefficients learning and sparse regularization. Section IV presents the details of the approach investigated to cope with the final optimization problem. Section V discusses the experimental results. Section VI concludes the paper.

II. Problem Formulation

A high spatial, high spectral image, $X = [x_1, \ldots, x_n] \in \mathbb{R}^{m_\lambda \times n}$ (target image), is recovered by fusing a high spectral HSI, $Y_H \in \mathbb{R}^{m_\lambda \times m}$, and a high spatial resolution MSI, $Y_M \in \mathbb{R}^{n_\lambda \times n}$, where m_λ, n_λ $(n_\lambda < m_\lambda)$ are the number of bands of HSI and MSI, respectively, and m, n are the number of pixels in each band of HSI and MSI, respectively. The HSI is regarded as a downsampled, blurred version of the target image; the MSI is a spectral degradation of the target image. The corresponding degradation model is expressed as follows: [39], [40]

$$Y_H = XBS + N_H, Y_M = RX + N_M \tag{1}$$

where X is the full resolution target image, Y_H and Y_M are the observed HSIs and MSIs, respectively. $B \in \mathbb{R}^{n \times n}$ denotes a cyclic convolution operator acting on each band; $S \in \mathbb{R}^{n \times m}$ denotes a downsampling matrix; $R \in \mathbb{R}^{n_\lambda \times m_\lambda}$ represents the spectral response function of the MS sensor; N_H, N_M are the noise matrices of the HSI and MSI, respectively. In this paper, both N_H and N_M are drawn from a normal distribution.

According to the storage characteristics of the BIP (the band interleaved by pixels) data format, where each pixel is cross arranged in band order [45], the unknown target high spatial, high spectral image, X, is decomposed as $X = [x_1, \ldots, x_n]$, where $x_i = [x_{i,1}, x_{i,2}, \ldots, x_{i,m_\lambda}]^T$ represents the $m_\lambda \times 1$ vector corresponding to the ith spatial position. Because the adjacent bands of the HSI are generally spectrally correlated, the vector x_i of the HSI usually exists in a low-dimensional subspace whose dimension is much smaller than the number of bands, m_λ, in the HSI [46]. It can be further expressed as $x_i = Hu_i$, where u_i denotes the projection of x_i onto the subspace spanned by the column of $H^T \in \mathbb{R}^{m_\lambda \times \tilde{m}_\lambda}$. The value of H is obtained by defining *a priori* on the scene or directly determined from the HSI; H is an orthogonal matrix such that $H^T H = I_{\tilde{m}_\lambda}$. Therefore, the target image, X, is represented as follows:

$$X = HU \tag{2}$$

where $U = [u_1, \ldots, u_n]$ with $U \in \mathbb{R}^{\tilde{m}_\lambda \times n}$. H is invertible, thereby satisfying $U = H^T X$. In the proposed fusion framework, we refer to principal component analysis (PCA) [39] to calculate H from the HSI. PCA is used only to learn the required subspace (Algorithm 1, Step 1). Because the dimension of subspace $\tilde{m}_\lambda$ is much smaller than the number of HS bands, namely, $\tilde{m}_\lambda \ll m_\lambda$, fusing in the subspace greatly reduces the computational complexity of the fusion algorithm, while avoiding possible matrix singularity caused by the strong spectral correlation of the HSI.

Algorithm 1: Fusion of HSIs and MSIs based on Bayesian Nonparametric Sparse Representation.

Input: $Y_H, Y_M, N_H, N_M, B, S, R, \tilde{m}_\lambda K$;
% Determinate the subspace from Y_H
1 $\hat{H} \leftarrow PCA(Y_H, \tilde{m}_\lambda)$;
% Propose a rough estimation $\tilde{U}$ of U
2 Calculate $\tilde{U} \triangleq \hat{\mu}_{U|Y_M}$ by using the method in [47];
% Learning dictionary and sparse coefficients (see III-C)
3 **for** $i = 1$ **to** $\tilde{m}_\lambda$ **do**
4 $(\bar{D}_i, \bar{A}_i, \pi, \gamma, \gamma_s) \leftarrow \text{Beta} - \text{Bernoulli}(\tilde{U}_i)$;
5 **end**
% Alternate optimization
6 **for** $t = 1, 2, \ldots$ **to** *stopping rule 1* **do**
% Optimize with respect to U using SALSA
7 $\hat{U}^{(t)} \in \{U | L(U, \hat{A}^{(t-1)}) \leq L(\hat{U}^{(t-1)}, \hat{A}^{(t-1)})\}$;
% Update with respect to A (see IV-B)
8 $\hat{A}^{(t)} \leftarrow \text{Beta} - \text{Bernoulli}(\hat{U}^{(t-1)}, \bar{D}_i, \pi, \gamma, \gamma_s)$;
% Identify the sparse coefficient
9 Set $\bar{\Omega}_i = \{(j,k) \hat{A}^{(t)}(j,k) \neq 0\}$;
10 **end**
11 Recover $\hat{X} = \hat{H}\hat{U}$;
Output: $\hat{X}$

III. Method

A. Ill-Posed Inverse Problem

Using (2), (1) is reformulated as follows:

$$Y_H = HUBS + N_H, Y_M = RHU + N_M \tag{3}$$

It can be seen that recovering the high spatial resolution and high spectral resolution target image, X, from the observation images, Y_H and Y_M are linear inverse problems (LIPs) [44]. Whether the LIP problem is a well-posed problem or an ill-posed problem depends mainly on the dimension of the subspace and the number of spectral bands [40]. In this paper, we focus merely on the ill-posed case.

According to the statistical properties of the noise matrices, N_H and N_M, the observations, Y_H and Y_M, are assumed to be drawn from a normal distribution, expressed as follows:

$$\begin{aligned} Y_H &\sim \mathcal{N}_{m_\lambda, m}(HUBS, \Lambda_H, I_H) \\ Y_M &\sim \mathcal{N}_{n_\lambda, n}(RHU, \Lambda_M, I_M) \end{aligned} \tag{4}$$

Because the observations, Y_H and Y_M, are acquired by different (or heterogeneous) sensors, the noise matrices, N_H and N_M, are sensor-dependent and generally considered to be independent of each other. Therefore, in terms of the Bayesian theorem, the posterior distribution of U after taking the negative

logarithm on both sides is expressed as follows:

$$\begin{aligned}
&-\log p\left(U\left|Y_H, Y_M\right.\right) \\
&\geq -\log p\left(Y_H|U\right)-\log p\left(Y_M|U\right)-\log p\left(U\right) \\
&=\frac{1}{2}\left\|\Lambda_H^{-\frac{1}{2}}\left(Y_H-HUBS\right)\right\|_F^2+\frac{1}{2}\left\|\Lambda_M^{-\frac{1}{2}}\left(Y_M-RHU\right)\right\|_F^2 \\
&+\lambda\phi\left(U\right) \qquad (5)
\end{aligned}$$

where $\|X\|_F$ denotes the Frobenius norm of *X*. The first two terms represent the corresponding data fidelity terms with respect to HSIs and MSIs, respectively. $\phi(U)$ is the regularization term, and λ is the regularization parameter, whose optimal value is affected by the noise level. In this paper, the main purpose is to cope with the image fusion problem by calculating the maximum *a posteriori* estimator of *U*, which is equivalent to minimizing (5).

B. Sparse Representation of the Regularization Terms

In signal processing literature, sparse representation has received much attention, because of the use of the self-similarity property in terms of natural images [32], [40]. Because of its sparsity, sparse representation is generally used to regularize the fusion problem [37], [40]. As seen from the definition, the reconstructed signal can be approximated by the product of a prepared overcomplete dictionary and sparse coefficients [32], which means that the image patches of the target image projected onto a low-dimensional subspace can be represented by a linear combination of an appropriate overcomplete dictionary and sparse coefficients. The overcomplete dictionary can be a well-trained dictionary or be tuned to the input images. In this paper, we adopt the Bayesian nonparametric approach with beta process prior [41] to learn the dictionary from the input images, Y_H and Y_M. The regularization term is expressed as follows:

$$\phi\left(U\right)=\frac{1}{2}\sum_{i=1}^{\tilde{m}_\lambda}\left\|U_i-\mathcal{P}\left(\bar{D}_i\bar{A}_i\right)\right\|_F^2 \qquad (6)$$

where $U_i \in \mathbb{R}^n$ is the *i*th $(i=1,2,\ldots,\tilde{m}_\lambda)$ band of $U \in \mathbb{R}^{\tilde{m}_\lambda \times n}$, and $\bar{D}_i \in \mathbb{R}^{n_p \times K}$ is an overcomplete dictionary obtained from the input images. $\bar{A}_i \in \mathbb{R}^{K \times n_{\mathrm{pat}}}$, containing a small number of nonzero elements, are the sparse coefficients associated with the *i*th band. $\mathcal{P}(\cdot): \mathbb{R}^{n_p \times n_{\mathrm{pat}}} \mapsto \mathbb{R}^{n \times 1}$ represents a linear operator that averages the overlapping image patches of each band. The size of overlapping patches is $\sqrt{n_p} \times \sqrt{n_p}$; the adjoining operation is denoted as $\mathcal{P}^*(\cdot): \mathbb{R}^{n \times 1} \mapsto \mathbb{R}^{n_p \times n_{\mathrm{pat}}}$, such that $\mathcal{P}[\mathcal{P}^*(X)]=X$.

C. Bayesian Nonparametric Dictionary Learning

Referring to [47], we first construct the rough estimation, $\tilde{U}$ of *U*, from the observations, Y_H and Y_M (Algorithm 1, Step 2) and then use the Bayesian nonparametric dictionary learning method [41], [42], [48], to learn the overcomplete dictionary, $\bar{D}_i$, and sparse coefficients, $\bar{A}_i$, associated with the band, $\tilde{U}_i$ of $\tilde{U}$ (Algorithm 1, Step 4). Regarding $\tilde{U}_i$ as trained samples, a hierarchical model, based on the beta-Bernoulli process [48], [49] dictionary learning, is written as follows:

$$\begin{aligned}
&\tilde{U}_i=\bar{D}_i\bar{A}_i+\varepsilon_i \ \forall i \in \{1,\ldots,n_pK\} \\
&d_k \sim \mathcal{N}(0,\upsilon^{-1}I_\upsilon) \ \forall k \in \{1,\ldots,K\} \\
&\bar{A}_i=\mathbf{s}_i \odot \mathbf{z}_i \ \ s_{ik} \sim \mathcal{N}\left(\hat{\mu}_{ik_0},\widehat{\sum}_{ik_0}\right) \ \ \varepsilon_i \sim \mathcal{N}\left(0,\gamma^{-1}I_\upsilon\right) \\
&z_{ik} \sim \mathrm{Bern}(\pi_{k_0}) \ \pi_k \sim \mathrm{Beta}(\tau_0\eta_0,\tau_0(1-\eta_0)) \\
&\gamma_s \sim \mathrm{Gamma}(e_0,f_0) \ \gamma \sim \mathrm{Gamma}(c_0,d_0) \qquad (7)
\end{aligned}$$

In the above model, $\odot$ denotes the element-wise product; d_k is the dictionary element of $\bar{\mathrm{D}}_i$; υ represents the dimensionality of the dictionary elements. I_υ denotes the $\upsilon \times \upsilon$ identity matrix; *K* is the dictionary size. Bern, Beta, and Gamma denote Bernoulli, Beta, and Gamma distributions, respectively. s_{ik} are the sparse valued weights; z_{ik} are the binary valued assignments to record whether the dictionary element is activated for the corresponding observations. Conjugate beta prior is placed over π_k with hypermeters τ_0 and η_0, which denotes the probability of using element d_k; γ_s and γ represent the precision of the sparse weights and observation noise, respectively; c_0, d_0, e_0, and f_0 denote the hyperparameters; ε_i is the observation noise. The above hierarchical model is completely conjugate; hence the Gibbs sampling approach [43] can be used over it for Bayesian inference. To acquire the maximum posterior, we use the Gibbs sampling method to iteratively sample from the conditional distribution of each latent variable given the others and the input data. This defines a Markov chain Monte Carlo (MCMC) whose stationary distribution is the posterior distribution [50]. The corresponding sampling formulas for the Gibbs sampling process used in our method are as follows:

Sample d_k: from $\mathcal{N}(\tilde{\mu}_k,\widetilde{\sum}_k)$, where

$$\widetilde{\sum}_k=\left(2\upsilon I_{2\upsilon}+\gamma\sum_{i=1}^{N}\left(s_{ik}z_{ik}\right)^2\right)^{-1} \tilde{\mu}_k=\gamma\widetilde{\sum}_k\sum_{i=1}^{N}s_{ik}\tilde{x}_{i(-k)} \qquad (8)$$

Sample z_{ik}: from $\mathrm{Bern}(\frac{\varsigma\pi_{k_0}}{1-\pi_{k_0}+\varsigma\pi_{k_0}})$, where

$$\varsigma=\exp\left(-\gamma\left(s_{ik}^2d_k^Td_k-2s_{ik}d_k^T\tilde{x}_{i(-k)}\right)/2\right) \qquad (9)$$

Sample s_{ik}: from $\mathcal{N}(\hat{\mu}_{ik},\widehat{\sum}_{ik})$, where

$$\hat{\sum}_{ik}=\left(\gamma_s+\gamma d_k^Td_k\right)^{-1},\hat{\mu}_{ik}=\gamma\hat{\sum}_{ik}d_k^T\tilde{x}_{i(-k)} \qquad (10)$$

Sample π_k: from $\mathrm{Beta}(a,b)$, where

$$a=\tau_0\eta_0+\sum_{i=1}^{N}z_{ik},b=N-\sum_{i=1}^{N}z_{ik}+\tau_0\left(1-\eta_0\right) \qquad (11)$$

Sample γ: from $\mathrm{Gamma}(c,d)$, where

$$c=c_0+N\upsilon,d=d_0+1/2\sum_{i=1}^{N}\left\|x_i-\bar{D}_i\left(\mathbf{s}_i\odot\mathbf{z}_i\right)\right\|_2^2 \qquad (12)$$

Sample γ_s: from $\mathrm{Gamma}(e, f)$, where

$$e = e_0 + NK/2, f = f_0 + 1/2 \sum_{i=1}^{N} \|\mathbf{s}_i\|_2^2 \tag{13}$$

where $\tilde{x}_{i(-k)} = x_i - \bar{D}_i(\mathbf{s}_i \odot \mathbf{z}_i) + d_k(\mathbf{s}_{ik} \odot \mathbf{z}_{ik})$ represents the reconstruction error using all but the *k*th dictionary element. As a result of MCMC inferencing, we obtain the posterior distributions of the five latent variables.

D. Sparse Coding for the Regularization Term

We assume that the prior distributions of the images U_i $(i = 1, \ldots, \tilde{m}_\lambda)$ of U are independent, and the marginal distribution $p(U_i)$ of U_i is drawn from a normal distribution with mean $\mathcal{P}(\bar{D}_i \bar{A}_i)$. Besides, as seen in Section III-C, $\bar{A}_i$ is a binary vector that encodes which $\bar{D}_i$, varying depending upon the update of U_i, is activated for the estimated image U_i. Therefore, we adopt the strategy that jointly updates U_i and $\bar{A}_i$ with $\bar{D}_i$ fixed to minimize the regularization term, $\phi(U)$. Using the joint prior distribution of U and *A*, (6) is reduced to the following:

$$\phi(U, A) = 1/2 \sum_{i=1}^{\tilde{m}_\lambda} \left\|U_i - \mathcal{P}\left(\bar{D}_i \bar{A}_i\right)\right\|_F^2 = 1/2 \left\|U - \bar{U}\right\|_F^2$$
$$\text{s.t. } A_i \odot M_i = 0 \tag{14}$$

where $\bar{U} = \mathcal{P}(\bar{D}_i \bar{A}_i)$ with $i = 1, \ldots, \tilde{m}_\lambda$. M_i is a matrix consisting of 0 and 1, which satisfies $M_i(j,k) = 1$ when $A_i(j,k) = 0$; otherwise, $M_i(j,k) = 0$. All zero elements in A_i marked through the label matrix M_i determined by constraints avoid the case where the structure of A_i is destructed by nonzero elements during alternate optimization. Besides, the position set of the nonzero elements of A_i, namely, $\bar{\Omega}_i = \{(j,k)|A_i(j,k) \neq 0\}$ are also identified (Algorithm 1, Step 9). We only need to calculate the nonzero elements during the update of A_i, which results in a much more computationally efficient approach.

Combining (14) and (5), the final optimization problem is written as follows:

$$\min_{U,A} L(U, A) \triangleq \frac{1}{2} \left\|\Lambda_H^{-\frac{1}{2}} (Y_H - HUBS)\right\|_F^2$$
$$+ \frac{1}{2} \left\|\Lambda_M^{-\frac{1}{2}} (Y_M - RHU)\right\|_F^2 + \frac{\lambda}{2} \left\|U - \bar{U}\right\|_F^2$$
$$\text{s.t. } \{A_i \odot M_i = 0\}_{i=1}^{\tilde{m}_\lambda} \tag{15}$$

IV. Alternate Optimization

As shown in (15), the optimization problem is a standard quadratic constrained optimization problem with respect to *U* and *A*. Because the computational dimension among matrices is large and the related operators are difficult to diagonalize, such a constrained optimization problem is usually difficult to solve. Hence, as in the case of Wei *et al.* [40], to cope with this problem, we implement the strategy that holds *A* fixed to update *U*, and then updates *A* with *U* fixed.

The ADMM technique, elaborated on by Wei *et al.* [40], is used to optimize *U*, whose convergence is guaranteed by the Eckstein–Bertsekas theorem [51]. For optimization in terms of A, we apply the Bayesian nonparametric approach described in Section III-C. It is noteworthy that the posterior distributions of $\bar{D}_i$, π, γ, γ_s, obtained in Section III-C, are served *a priori* at this stage; only $\bar{A}_i$, γ, and γ_s are updated with Gibbs sampling.

A. Updating U With ADMM

Holding *A* fixed (equivalent to fixing $\bar{U}$), the optimization problem with respect to *U* in (15) is translated into the following:

$$\min_U L(U) \triangleq \frac{1}{2} \left\|\Lambda_H^{-\frac{1}{2}} (Y_H - HUBS)\right\|_F^2$$
$$+ \frac{1}{2} \left\|\Lambda_M^{-\frac{1}{2}} (Y_M - RHU)\right\|_F^2 + \frac{\lambda}{2} \left\|U - \bar{U}\right\|_F^2 \tag{16}$$

By adding the variables $V_1 = UB$, $V_2 = U$, and $V_3 = U$, the above formula is converted into the following form associated with the augment Lagrangian approach:

$$L(U, V_1, V_2, V_3, G_1, G_2, G_3) = \frac{1}{2} \left\|\Lambda_H^{-\frac{1}{2}} (Y_H - HV_1S)\right\|_F^2$$
$$+ \frac{\mu}{2} \left\|UB - V_1 - G_1\right\|_F^2 + \frac{1}{2} \left\|\Lambda_M^{-\frac{1}{2}} (Y_M - RHV_2)\right\|_F^2$$
$$+ \frac{\mu}{2} \left\|U - V_2 - G_2\right\|_F^2 + \frac{1}{2} \left\|\bar{U} - V_3\right\|_F^2 + \frac{\mu}{2} \left\|U - V_3 - G_3\right\|_F^2 \tag{17}$$

where G_1, G_2, and G_3 are the Lagrangian multipliers ($\mu > 0$). The updates of *U*, V_1, V_2, V_3, G_1, G_2, and G_3 are achieved by the SALSA approach [40], [44].

B. Bayesian Nonparametric Sparse Coding

The above section implements the optimization with respect to *U*. Holding *U* fixed, the optimization problem with respect to the sparse coefficients *A* (equivalent to optimizing $\bar{U}$) is converted as follows:

$$\hat{A}_i = \arg\min_{A_i} \left\|p_i - \bar{D}_i A_i\right\|_F^2, \text{s.t. } \{A_i \odot M_i = 0\}_{i=1}^{\tilde{m}_\lambda} \tag{18}$$

where $p_i = \mathcal{P}^*(U_i)$. As described in Section III-D, we optimize only the nonzero elements in A_i, which are easily solved.

In this section, to solve the optimization problem, we continue to use the Bayesian nonparametric approach with the beta-Bernoulli process. The solution is basically the same as that described in Section III-C. The main differences are that 1) the posterior distributions of $\bar{D}_i$, π, γ, γ_s, obtained from Section III-C, are used as the initial values in this phase; 2) fixing dictionary, $\bar{D}_i$, and the usage probability of dictionary elements, π, the optimization with respect to *A* is implemented by using the corresponding Gibbs sampling equations to update the components s_{ik} and z_{ik}.

C. Complexity Analysis

Compared with optimization methods, the MCMC method is computationally costly. The computational complexity of the proposed approach consists mainly of three parts: 1) Bayesian nonparametric dictionary learning; 2) optimization with respect to *U* by using the SALSA algorithm; 3) sparse coding with respect to A. The complexity of the Bayesian nonparametric

dictionary learning is $\mathcal{O}(KT\tilde{m}_\lambda \log Kn_p n_{\text{pat}})$, where *T* is the number of Gibbs sampling iterations. The complexity of the SALSA algorithm is $\mathcal{O}(n_{it}\tilde{m}_\lambda n \log(\tilde{m}_\lambda n))$ [40], where n_{it} is the number of SALSA iterations. The complexity of the sparse coding with respect to *A* is $\mathcal{O}(KH\tilde{m}_\lambda \log Kn_p n_{\text{pat}})$, where *H* is the number of sparse coding iterations.

D. Experimental Datasets

The Reflective Optics System Imaging Spectrometer (ROSIS) dataset, acquired by the ROSIS sensor in 2003 over Pavia, Italy, has 610 × 340 pixels with a spatial resolution of 1.3 m per pixel. The original ROSIS image contains 115 spectral bands covering the spectral range from 0.430 to 0.838 μm, which was widely used in [40], [52], [53]. After the 22 bands corrupted by water absorption and noisy effects have been discarded, the remaining 93 bands are adopted for analysis.

The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) dataset, gathered by the AVIRIS sensor over Indian Pine in 1996, contains 512 × 614 pixels with a spatial resolution of 20 m, and has 224 bands covering the spectral range from 400 to 2500 nm [54].

The New York dataset, consisting of HSIs and MSIs that were collected by two different sensors over the same zone. The HSI was collected by the Hyperion sensor, which provides 242 bands from 0.4 to 2.5 μm, with 30 m of spatial resolution. The MSI was captured by the Advanced Land Imager (AL1) sensor, which provides nine bands with 30 m of spatial resolution, covering the 0.433–0.453, 0.45–0.515, 0.525–0.605, 0.63–0.69, 0.775–0.805, 0.845–0.89, 1.2–1.3, 1.55–1.75, and 2.08–2.35 μm regions of the electromagnetic spectrum. Both sensors were carried by the EO-1 satellite [55]–[57].

E. Fusion Quality Metrics

To evaluate the quality of the proposed fusion approach, the following six widely used and complementary quality measures are investigated in our study. Referring to [6] and [39], we propose to use the restored SNR (RSNR), root mean square error (RMSE), spectral angle mapper (SAM), universal image quality index (UIQI), relative dimensionless global error in synthesis (ERGAS) [58], and degree of distortion (DD) as quantitative measures, which are defined as follows:

1) *RSNR:* RSNR is used to reflect the difference between the reference image, X, and the fused image,$\hat{X}$. The larger the RSNR, the better the fusion quality and vice versa. RSNR is defined as follows:

$$\text{RSNR}\left(X,\hat{X}\right) = 10\log_{10}\left(\|X\|^2 / \left\|X-\hat{X}\right\|_2^2\right) \quad (19)$$

2) *RMSE:* RMSE measures the similarity between the reference image, X, and the fused image $\hat{X}$. The smaller the RMSE, the better the fusion quality. RMSE is defined as follows:

$$\text{RMSE}\left(X,\hat{X}\right) = 1/nm_\lambda \left\|X-\hat{X}\right\|_F^2 \quad (20)$$

3) *SAM:* SAM is commonly used to quantify the spectral information preservation. SAM is defined as follows:

$$\text{SAM}\left(x_n,\hat{x}_n\right) = \arccos\left(\langle x_n,\hat{x}_n\rangle / \|x_n\|_2\|\hat{x}_n\|_2\right) \quad (21)$$

where x_n and $\hat{x}_n$ represent the spectral vector of x and $\hat{x}$, respectively. We use the average SAM value calculated between the spectral vectors of the reference image and the estimated image. The smaller the SAM, the less the spectral distortion.

4) *UIQI:* UIQI reflects the similarity between the single-band images of the reference image and the fused image, namely, $a = [a_1,\ldots,a_N]$ and $\hat{a} = [\hat{a}_1,\ldots,\hat{a}_N]$. The larger the UIQI, the smaller the spectral distortion. The value is related to the correlation, luminance, and contrast distortions of the fused image with respect to the reference image. UIQI is defined as follows:

$$\text{UIQI}\left(a,\hat{a}\right) = 4\sigma^2_{a\hat{a}}\mu_a\mu_{\hat{a}} / \left(\sigma_a^2+\sigma_{\hat{a}}^2\right)\left(\mu_a^2+\mu_{\hat{a}}^2\right) \quad (22)$$

where $(\mu_a,\mu_{\hat{a}},\sigma_a^2,\sigma_{\hat{a}}^2)$ represents the means and variances of a and $\hat{a}$, and $\sigma^2_{a\hat{a}}$ is the covariance of $(a,\hat{a})$.

5) *ERGAS:* ERGAS calculates the amount of spectral distortion and provides a global statistical measure of the fused image with the best value at zero. ERGAS is defined as follows:

$$\text{ERGAS} = 100\times m/n\sqrt{1/m_\lambda\sum_{i=1}^{m_\lambda}\left(\text{RMSE}\left(i\right)/\mu_i\right)^2} \quad (23)$$

where μ_i is the mean of the *i*th band of the HSI; m_λ is the number of the HS bands; and m/n is the ratio between the HSIs and MSIs. The smaller the ERGAS, the less the spectral distortion.

6) *DD:* DD reflects the degree of spectral distortion of the reference image and the fused image. The smaller the value, the higher the fusion quality. The measure of fusion quality is defined as follows:

$$\text{DD}\left(X,\hat{X}\right) = 1/nm_\lambda\left\|\text{vec}\left(X\right)-\text{vec}\left(\hat{X}\right)\right\|_1 \quad (24)$$

where $\text{vec}(\cdot)$ represents the vectorization of the matrix.

V. Results and Discussion

A. Experiments With ROSIS Data

We propose reconstructing the reference image, *X*, from the low spatial resolution, HSI, and the high spatial resolution, MSI, by using the proposed algorithm.

For simulation experiments, we selected, from the ROSIS dataset, three scenes (Scenes 1–3) with dimensions of 128 × 128 × 93. The related composite color images, formed by selecting the red, green, and blue bands of the reference images with respect to Scenes 1–3, are depicted in Fig. 1(a)–(c).

In all three scenes, the reference images, *X*, are used to generate the degraded images, Y_H and Y_M. Specifically, the HSIs Y_H are generated by applying a 5 × 5 Gaussian spatial filter and by downsampling every 4 pixels in both horizontal and vertical directions for each band of the reference images, *X*.

Fig. 1. Reference images with respect to Scenes 1–3 of ROSIS dataset. (a) Scene 1. (b) Scene 2. (c) Scene 3. (d) IKONOS-like spectral responses.

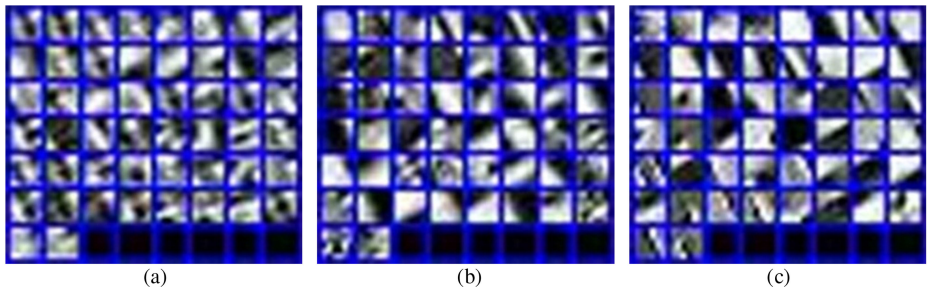

Fig. 2. Dictionaries of the first three bands with respect to Scene 1 of ROSIS data. (a) Dictionary for Band 1. (b) Dictionary for Band 2. (c) Dictionary for Band 3.

Besides, four-band MSIs Y_M are obtained by filtering the reference images, X, with the IKONOS-like spectral response shown in Fig. 1(d). Providing for reality, the HSIs and MSIs are contaminated by zero-mean additive Gaussian noise, and the noise power $s^2_{H,i}$ and $s^2_{M,j}$ depend on the SNR defined by

$$\mathrm{SNR}_{H,i} = 10\log\left(\|(XBS)\|_F^2 / s^2_{H,i}\right) \tag{25}$$

$$\mathrm{SNR}_{M,i} = 10\log\left(\|(RX)\|_F^2 / s^2_{M,i}\right) \tag{26}$$

In all three scenes, we assumed that $\mathrm{SNR}_{H,i}$ is 35 dB for the first 43 bands, and $\mathrm{SNR}_{H,i}$ is 30 dB for the remaining 50 bands of the HSIs Y_H; whereas $\mathrm{SNR}_{M,i}$ of the MSIs Y_M is 30 dB for all bands.

B. Low-Dimensional Subspace Learning

We used the PCA similar to [39], to learn the low-dimensional subspace transform matrix, H. Note that the PCA, which is widely used in HSIs [38], [40], projects the original image onto a low-dimensional subspace while retaining most of the information. The experiment with respect to Scene 1 reveals that, when there are five bands in the subspace, $\tilde{m}_\lambda$, the eigenvectors contain nearly 99.9% of the spatial and spectral information of the original image. Therefore, in all three scenes, the first five eigenvectors of the HSIs are selected to build the subspaces of interest.

C. Dictionary Learning

As explained earlier, dictionary learning plays a vital role in the reconstruction of high spatial resolution HSIs. We use the Bayesian nonparametric approach with beta process [41] prior to learning the overcomplete dictionaries.

With our experiments in all three scenes, the input images, $\tilde{U}$, are divided into 8×8 pixel image patches for dictionary learning, with maximum overlap between adjacent patches. The dictionary elements and other parameters are all initialized with the singular value decomposition SVD method. The hyperparameters are $c_0 = d_0 = e_0 = f_0 = 10^{-6}$, $\tau_0 = 2$, and $\eta_0 = 0.5$. These are standard uninformative priors applied in, e.g., [48]. We initialized the Gibbs sampling with dictionary size $K = 50$, which is the value based on our prior belief that the total number of endmembers in a given image is generally less than fifty [36]. The number of iterations is 500, and Gibbs sampling is used to obtain the posterior distributions of the five latent variables. Especially noteworthy is the fact that the dictionary size, K, is inferred by the Bayesian nonparametric method through searching and deleting the unused dictionary elements during the Gibbs iterative sampling, which reduces the dictionary dimension and improves learning efficiency. Because dictionary learning is carried out in subspace, where the number of samples is relatively smaller, we do not infer the dictionary size nonparametrically. Fig. 2 shows the dictionaries of the first three bands, with respect to Scene 1 of the ROSIS data, in descending order according to the usage probability of the dictionary elements, π. As seen in Fig. 2, the learned dictionary elements well represent the structure and texture features of the target image.

D. Selection of the Regularization Parameter

To select an appropriate value for λ, we evaluate our proposed method as a function of λ. As shown in Fig. 3 the results

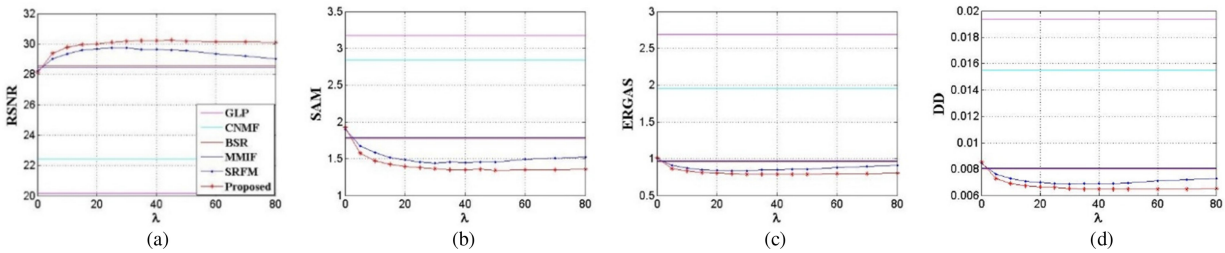

Fig. 3. Performance of the proposed method verse λ with respect to Scene 1 of ROSIS data. (a) RSNR. (b) SAM. (c) ERGAS. (d) DD.

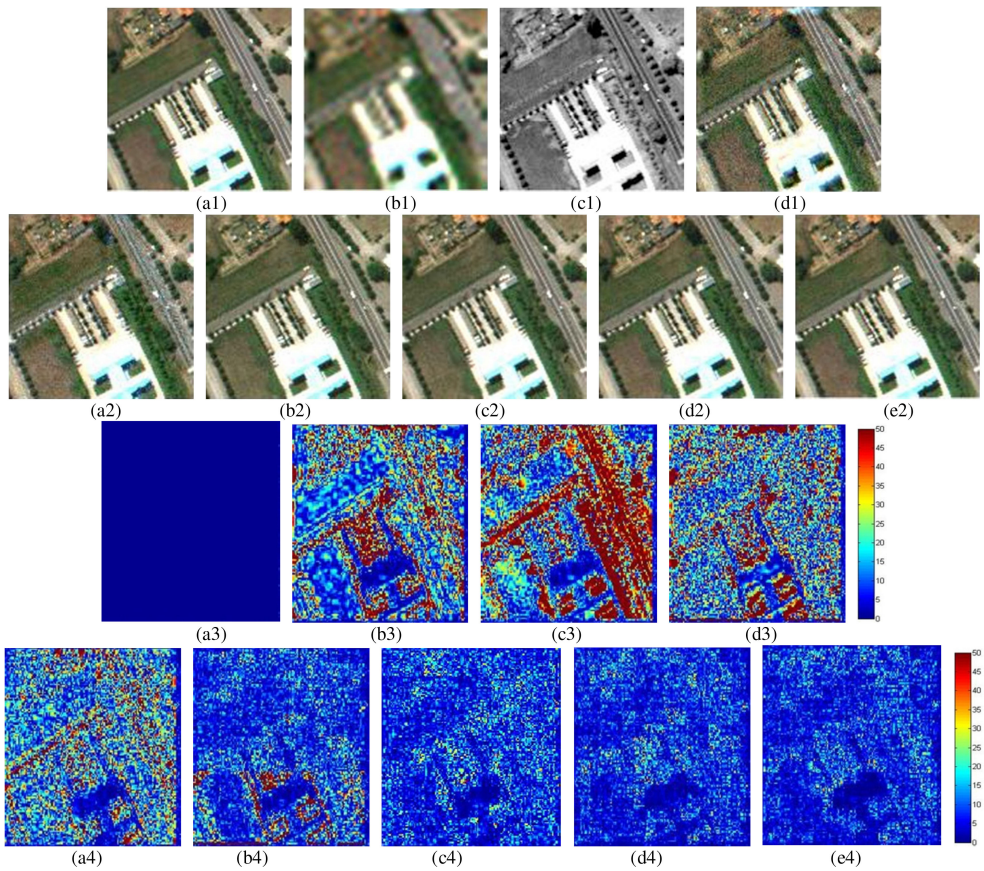

Fig. 4. Fusion results and absolute error images with respect to Scene 1 of ROSIS data. (a1) to (d1) and (a2) to (e2) reference, HSI, MSI, and fusion results using different methods, respectively. (a3) to (d3) and (a4) to (e4) absolute error images of the competing methods. (a1) Reference image X. (b1) HSI Y_H. (c1) MSI Y_M. (d1) GLP [10]. (a2) CNMF [22]. (b2) BSR [36]. (c2) MMIF [27]. (d2) SRFM [40]. (e2) Our method. (a3) Error of X. (b3) Error of Y_H. (c3) Error of Y_M. (d3) GLP [10]. (a4) CNMF [22]. (b4) BSR [36]. (c4) MMIF [27]. (d4) SRFM [40]. (e4) Our method.

with respect to Scene 1 of the ROSIS data that the values of the evaluation indexes tend to be stable when $\lambda \geq 35$, and the performance of our proposed algorithm is superior to that of the other five methods in the range of the given values of λ. Hence, for all three scenes, we choose $\lambda = 35$ to conduct the fusion experiments.

E. Comparative Study

We compared our approach with the other five approaches, which constitute the state-of-the-art in this area: GLP [10], CNMF [22], BSR [36], MMIF [27], and SRFM [40]. All the related experimental parameters, with respect to the five methods, are referenced to the original literature.

All the algorithms were implemented using MATLABR2012A on a computer with Intel(R) Xeon(R) CPU E3-1230 V2 @ 3.30 GHz and 8 GB RAM. The fusion results and absolute error images, with respect to Scene 1 of the ROSIS data, are depicted in Fig. 4. The performance of the proposed algorithm, while slightly better than that of the SRFM [40] algorithm, especially in the white ground area, is significantly superior to that of the GLP [10], CNMF [22], BSR [36], and MMIF [27] algorithms. Figs. 5 and 6 display the fusion results and absolute error images, with respect to Scenes 2 and 3 of the

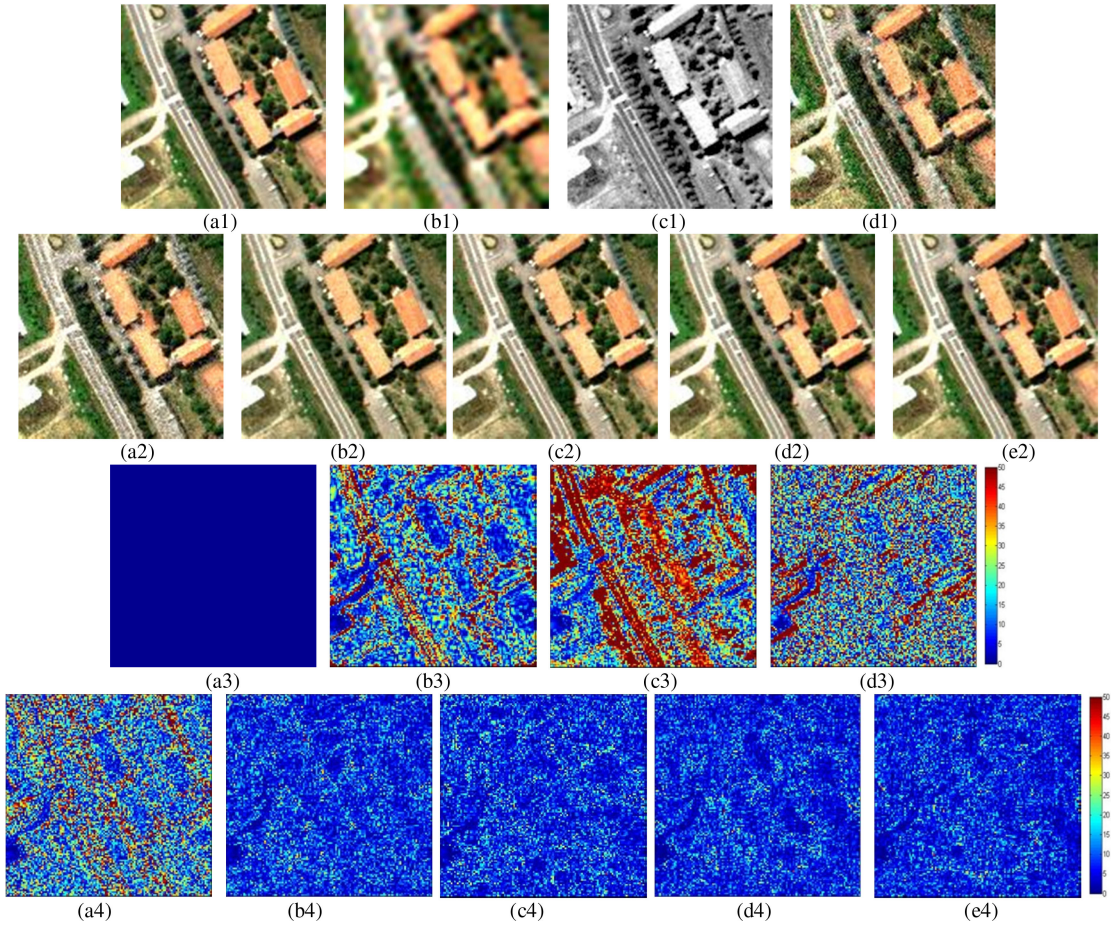

Fig. 5. Fusion results and absolute error images with respect to Scene 2 of ROSIS data. (a1) to (d1) and (a2) to (e2) reference, HSI, MSI, and fusion results using different methods, respectively. (a3) to (d3) and (a4) to (e4) absolute error images of the competing methods. (a1) Reference image X. (b1) HSI Y_H. (c1) MSI Y_M. (d1) GLP [10]. (a2) CNMF [22]. (b2) BSR [36]. (c2)MMIF [27]. (d2)SRFM [40]. (e2) Our method. (a3) Error of X. (b3) Error of Y_H. (c3) Error of Y_M. (d3) GLP [10]. (a4) CNMF [22]. (b4) BSR [36]. (c4) MMIF [27]. (d4) SRFM [40]. (e4) Our method.

ROSIS data, respectively, which show the same performance of the proposed algorithm as in Scene 1. The main reason is that using the Bayesian nonparametric method to learn the dictionary elements and sparse coefficients and optimizing the sparse coefficients in the alternate optimization process introduces more *a priori* information to enhance the ability of the dictionary elements to express the structure and texture information of the target image, thereby further improving the quality of the fusion results. To illustrate more intuitively the difference in fusion quality under different algorithms, Tables I–III give the corresponding quantitative indicators with respect to the three scenes. As seen from the PSNR, RMSE, SAM, UIQI, ERGAS, and DD values, the performance of the algorithm is always optimal, which is consistent with the above visual results. However, compared with optimization methods, the MCMC method can be computationally costly; therefore, the computational time of the proposed method is relatively improved.

F. Experiments With AVIRIS Dataset

In this section, we analyze another group of experimental results of image fusion with AVIRIS data to further reveal the performance of the proposed algorithm. We select a $120 \times 120 \times 224$ AVIRIS image as the reference image. The simulation of Y_H is the same as that described in Section V-C, and the bands of Y_M correspond to Bands 1–5 and 7 of Landsat TM, whose wavelength ranges are 450–520, 520–600, 630–690, 760–900, 1550–1750, and 2080–2350 nm, respectively [59]. The spectral values of each band are the average of the spectral values in the corresponding band with respect to the reference image, including six bands. We contaminate Y_H and Y_M with zero-mean additive Gaussian noise, assuming that $\mathrm{SNR}_{H,i} = 300$ dB and $\mathrm{SNR}_{M,j} = 200$ dB, respectively. The proposed method is implemented to simulate HSI and MSI with a subspace of dimension $\tilde{m}_\lambda = 10$, which contains nearly 99.3% of the spatial and spectral information of the original image. We choose $\lambda = 40$ to conduct the fusion experiments. For dictionary training and sparse coefficient learning, the related parameters are the same as in Section V-C. Fig. 7 displays the reference images, HSI, MSI, and the fusion results under different algorithms. Additionally, the absolute error images of the fusion results are also given. Table IV shows the quantitative results corresponding to different algorithms. As shown in Fig. 7 and Table IV, the proposed fusion algorithm is still superior to GLP [10],

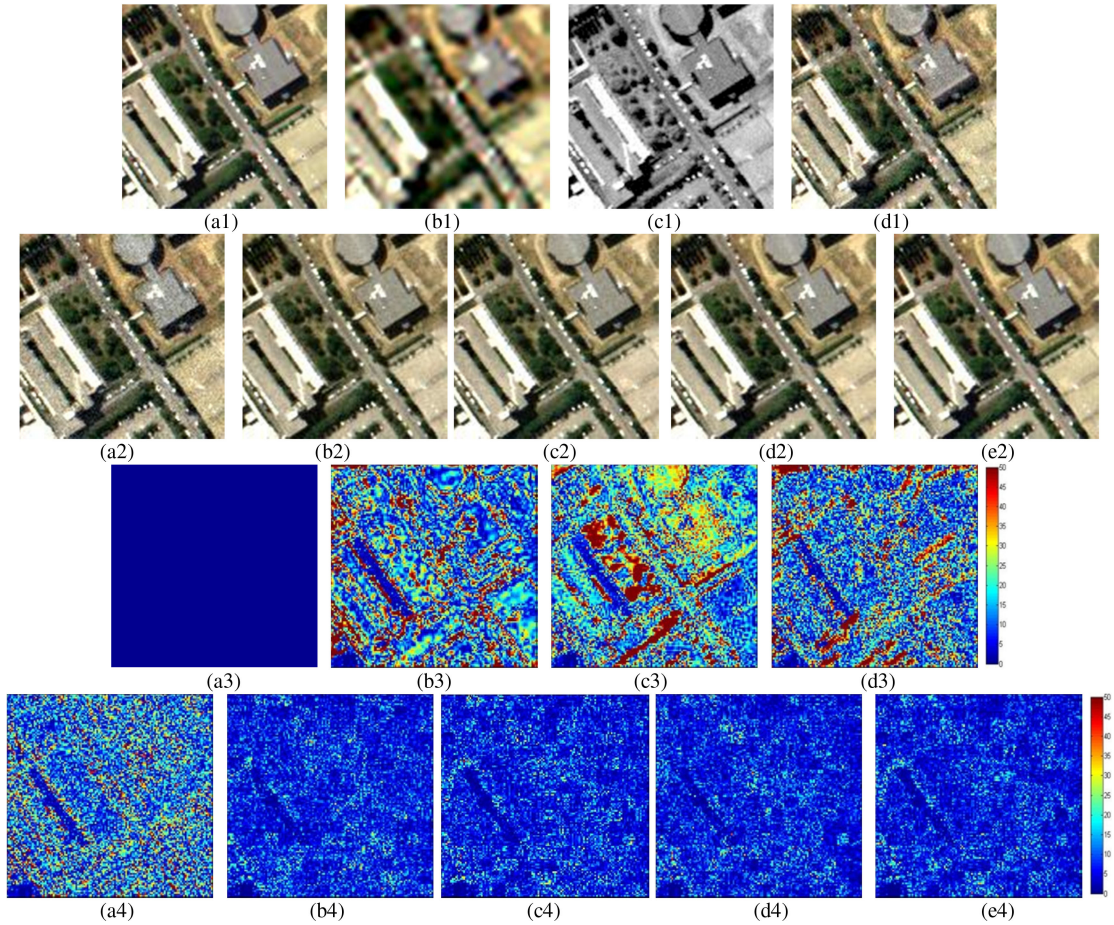

Fig. 6. Fusion results and absolute error images with respect to Scene 3 of ROSIS data. (a1) to (d1) and (a2) to (e2) reference, HSI, MSI, and fusion results using different methods, respectively. (a3) to (d3) and (a4) to (e4) absolute error images of the competing methods. (a1) Reference image X. (b1) HSI Y_H. (c1) MSI Y_M. (d1) GLP [10]. (a2) CNMF [22]. (b2) BSR [36]. (c2) MMIF [27]. (d2)SRFM [40]. (e2) Our method. (a3) Error of X. (b3) Error of Y_H. (c3) Error of Y_M. (d3) GLP [10]. (a4) CNMF [22]. (b4) BSR [36]. (c4) MMIF [27]. (d4) SRFM [40]. (e4) Our method.

TABLE I
PERFORMANCE OF DIFFERENT FUSION METHODS BASED ON SCENE I OF ROSIS DATA

Method	GLP [10]	CNMF [22]	BSR [36]	MMIF [27]	SRFM [40]	Ours
RSNR (dB)	20.162	22.445	28.551	28.427	29.752	**30.230**
RMSE (dB)	0.028	0.026	0.011	0.011	0.009	**0. 010**
SAM (°)	3.174	2.845	1.772	1.787	1.452	**1.356**
UIQI	0.926	0.957	0.989	0.989	0.992	**0.993**
ERGAS	2.687	1.956	0.962	0.973	0.836	**0.791**
DD ($\times 10^{-2}$)	1.938	1.552	0.799	0.807	0.688	**0.650**
Time (s)	10	**5**	173	62	11	380

TABLE II
PERFORMANCE OF DIFFERENT FUSION METHODS BASED ON SCENE II OF ROSIS DATA

Method	GLP [10]	CNMF [22]	BSR [36]	MMIF [27]	SRFM [40]	Ours
RSNR (dB)	19.363	20.696	28.814	28.647	30.106	**30.671**
RMSE (dB)	0.030	0.026	0.010	0.010	0.009	**0. 008**
SAM (°)	3.078	2.753	1.733	1.758	1.385	**1.313**
UIQI	0.887	0.916	0.986	0.986	0.990	**0.991**
ERGAS	2.852	2.339	0.930	0.942	0.794	**0.745**
DD ($\times 10^{-2}$)	2.227	1.854	0.787	0.790	0.007	**0.006**
Time (s)	10	**5**	220	74	10	442

CNMF [22], CSU [23], and SRFM [40] algorithms for AVIRIS data.

G. Experiments With New York Dataset

In order to evaluate the proposed algorithm under a more realistic scenario, the New York dataset has also been used. Since both the HSI and MSI have the same spatial resolution, the HSI has just been spatially degraded, performing the same simulation measures as that described in Section V-C. We make use of only the MSI bands 4, 7, and 9 to make the problem more challenging [55]. In addition, the images are co-registered before carrying on the fusion process, and the portion of 60×120 pixels has been selected for the fusion. The proposed method is implemented with a subspace of dimension $\tilde{m}_\lambda = 5$, which contains nearly

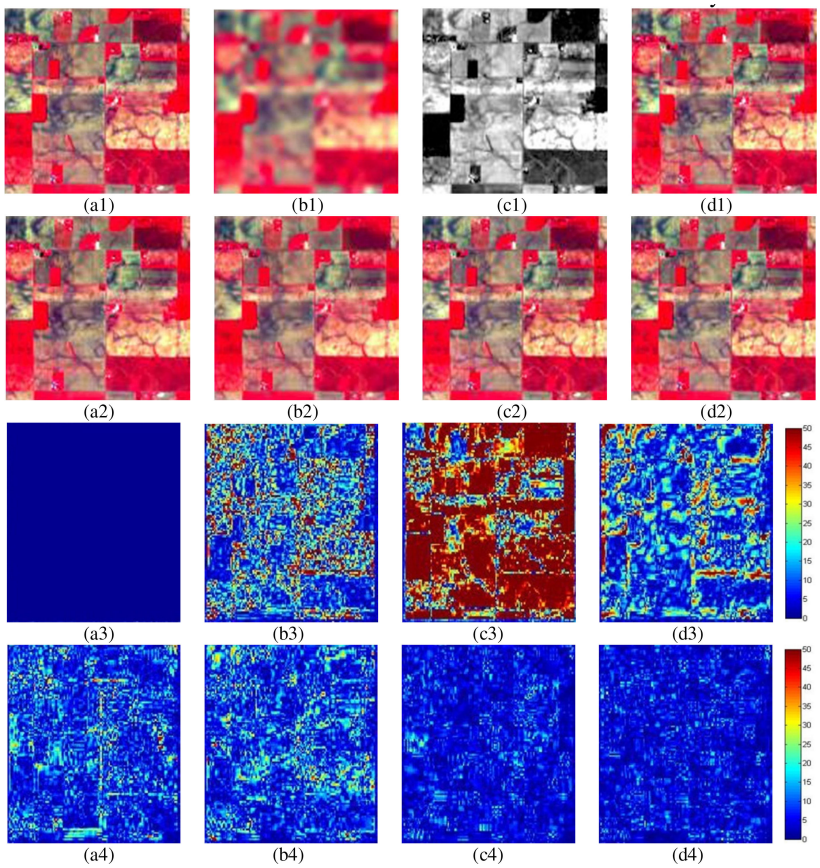

Fig. 7. Fusion results and absolute error images of AVRIS data. (a1) to (d1) and (a2) to (d2) reference, HSI, MSI, and fusion results using different methods, respectively. (a3) to (d3) and (a4) to (d4) absolute error images of the competing methods. (a1) Reference image X. (b1) HSI Y_H . (c1) MSI Y_M . (d1) GLP [10]. (a2) CNMF [22]. (b2) CSU [23]. (c2) SRFM [40]. (d2) Our method. (a3) Error of X. (b3) Error of Y_H . (c3) Error of Y_M . (d3) GLP [10]. (a4) CNMF [22]. (b4) CSU [23]. (c4) SRFM [40]. (d4) Our method.

TABLE III
PERFORMANCE OF DIFFERENT FUSION METHODS BASED ON SCENE III OF ROSIS DATA

Method	GLP [10]	CNMF [22]	BSR [36]	MMIF [27]	SRFM [40]	Ours
RSNR (dB)	19.241	19.564	27.327	26.939	28.171	**29.576**
RMSE (dB)	0.031	0.030	0.012	0.013	0.011	**0.009**
SAM (°)	2.925	3.163	1.816	1.849	1.546	**1.379**
UIQI	0.915	0.926	0.987	0.986	0.989	**0.992**
ERGAS	2.819	2.606	1.142	1.166	1.028	**0.865**
DD ($\times 10^{-2}$)	2.194	1.889	0.8	0.813	0.732	**0.652**
Time (s)	10	**6**	168	60	11	337

TABLE IV
PERFORMANCE OF DIFFERENT FUSION METHODS BASED ON AVRIS DATA

Method	GLP[10]	CNMF[22]	CSU[23]	SRFM[40]	Ours
RSNR (dB)	22.944	26.757	26.325	33.710	**34.683**
SAM (°)	2.614	1.670	1.815	1.080	**0.988**
UIQI	0.858	0.889	0.879	0.914	**0.917**
ERGAS	5.391	5.087	5.101	4.900	**4.862**
DD ($\times 10^{-2}$)	49.286	29.742	34.130	15.096	**13.422**
Time (s)	22	34	**16**	21	851

TABLE V
PERFORMANCE OF DIFFERENT FUSION METHODS BASED ON NEW YORK DATA

Method	GLP[10]	CNMF[22]	CSU[23]	SRFM[40]	Ours
RSNR (dB)	15.632	14.016	18.257	25.563	**25.959**
RMSE (dB)	0.014	0.017	0.010	0.005	**0.004**
SAM (°)	4.396	5.451	4.038	3.145	**2.958**
UIQI	0.684	0.616	0.684	0.723	**0.724**
ERGAS	3.423	4.040	2.764	1.707	**1.655**
DD ($\times 10^{-2}$)	0.740	0.867	0.523	0.305	**0.292**
Time (s)	5	**3**	7	8	160

99% of the spatial and spectral information of the original image, and the regularization parameter is set $\lambda = 35$. For dictionary training and sparse coefficient learning, the related parameters are the same as in Section V-C. Fig. 8 and Table V show the visual results and the corresponding quantitative results. It can be seen that the proposed fusion algorithm still outperforms the other algorithms for New York data. These results agree well with those we obtained with the abovementioned two datasets, proving that the proposed fusion algorithm improves the fusion quality.

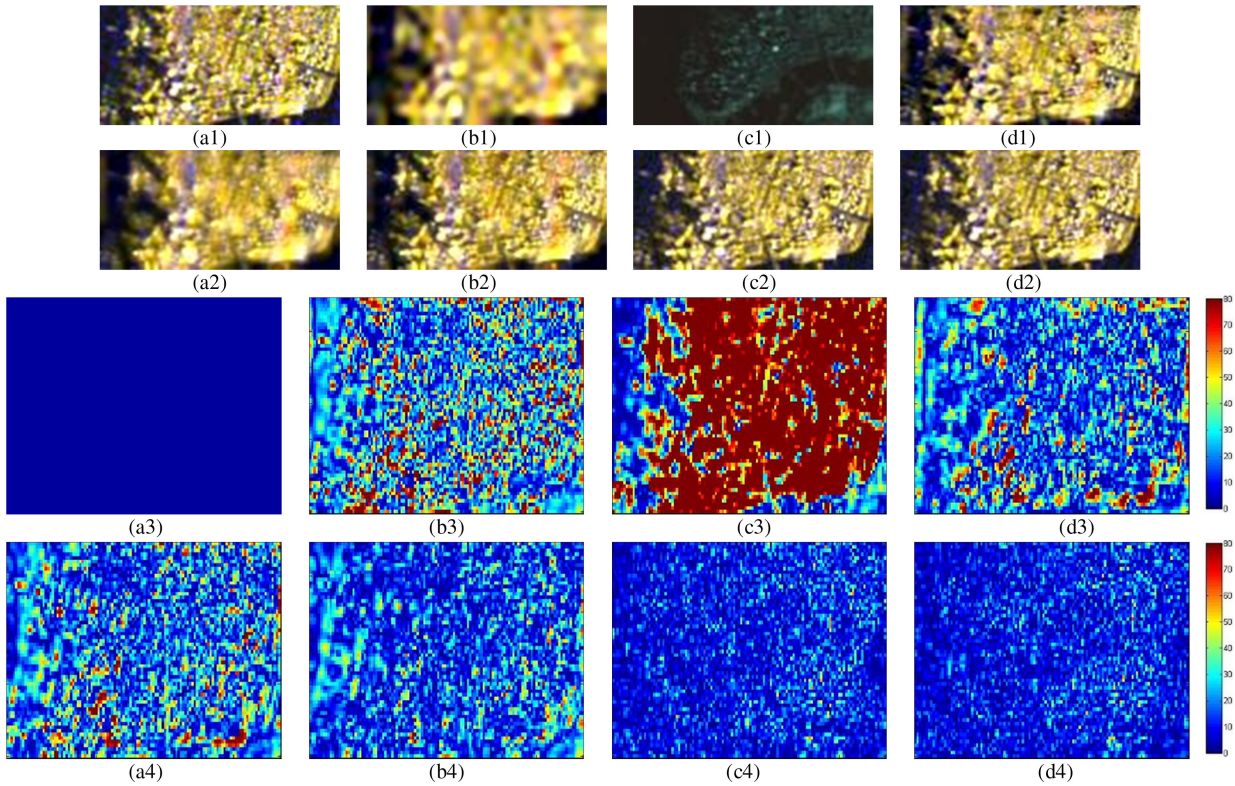

Fig. 8. Fusion results and absolute error images of New York data. (a1) to (d1) and (a2) to (d2) reference, HSI, MSI, and fusion results using different methods, respectively. (a3) to (d3) and (a4) to (d4) absolute error images of the competing methods. (a1) Reference image X. (b1) HSI Y_H . (c1) MSI Y_M . (d1) GLP [10]. (a2) CNMF [22]. (b2) CSU [23]. (c2) SRFM [40]. (d2) Our method. (a3) Error of X. (b3) Error of Y_H . (c3) Error of Y_M . (d3) GLP [10]. (a4) CNMF [22]. (b4) CSU [23]. (c4) SRFM [40]. (d4) Our method.

VI. Conclusion

In this paper, we proposed an effective HSI and MSI fusion method based on Bayesian nonparametric sparse representation. The proposed algorithm performs image fusion in subspace. Using the Bayesian nonparametric dictionary learning, the proposed algorithm learns the posterior distributions of the relevant parameters in the scenes of interest. Later, the information is applied to the alternate optimization algorithm, consisting of ADMM and the Bayesian nonparametric sparse coefficients learning to minimize the objective function. Exhaustive experiments with both two public datasets and one real-world dataset show that, with the advantages of offering smaller spatial structure error and smaller spectral distortion while obtaining superior quantitative results (except for the time index), the proposed algorithm outperforms the existing state-of-the-art. The main reason is that the Bayesian nonparametric method applied in this paper introduces more prior information to enhance the expression accuracy of the dictionary elements and the estimation precision of the target image. However, because the MCMC method used in the Bayesian model is computationally costly compared with optimization methods, more calculation time is required. Future works will jointly update the target image, sparse coefficients, and dictionary elements to achieve superior fusion results. Developing the parallel methods to improve the computational efficiency with respect to the Bayesian nonparametric model will also deserve some attention.

Acknowledgment

The authors would like to thank Dr. Q. Wei and Prof. N. Dobigeon for sharing the codes of [38]–[40], Dr. N. Yokoya for sharing the codes of [6], Dr. R. Arablouei for sharing the codes of [27], Prof. D. Landgrebe for providing the AVRIS Indian Pines data [54], and Prof. P. Gamba for providing the ROSIS Pavia data.

References

[1] L. Guanter *et al.*, "The EnMAP spaceborne imaging spectroscopy mission for earth observation," *Remote Sens.*, vol. 7, no. 7, pp. 8830–8857, 2015.
[2] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.
[3] A. Eckardt *et al.*, "DESIS (DLR Earth sensing imaging spectrometer for the ISS-MUSES platform)," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Milan, Italy, Jul. 2015, doi: 10.1109/IGARSS.2015.7326053.
[4] J. An, X. Zhang, H. Zhou, and L. Jiao, "Tensor-based low-rank graph with multimanifold regularization for dimensionality reduction of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4731–4746, Aug. 2018.
[5] Y. Yang, L. Wu, S. Y. Huang, Y. J. Tang, and W. G. Wan, "Pansharpening for multiband images with adaptive spectral–intensity modulation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 9, pp. 3196–3208, Sep. 2018.
[6] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 2, pp. 29–56, Jun. 2017.
[7] W. J. Carper and T. M. Lillesand, "The use of intensity-hue-saturation transform for merging SPOT panchromatic and multispectral image data," *Photogrammetric Eng. Remote Sens.*, vol. 56, no. 4, pp. 459–467, Jan. 1990.

[8] B. Aiazzi, S. Baronti, and M. Selva, "Improving component substitution pansharpening through multivariate regression of MS + Pan data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3230–3239, Oct. 2007.

[9] J. G. Liu, "Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details," *Int. J. Remote Sens.*, vol. 21, no. 18, pp. 3461–3472, 2000.

[10] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "MTFtailored multiscale fusion of high-resolution MS and PAN imagery," *Photogrammetric Eng. Remote Sens.*, vol. 72, no. 5, pp. 591–596, May 2006.

[11] X. X. Zhu, C. Grohnfeldt, and R. Bamler, "Exploiting joint sparsity for pansharpening: The J-SparseFI algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 2664–2681, May 2016.

[12] S. Li and B. Yang, "A new pan-sharpening method using a compressed sensing technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 738–746, Feb. 2011.

[13] X. X. Zhu and R. Bamler, "A sparse image fusion algorithm with application to pan-sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2827–2836, May 2013.

[14] L. Loncan *et al.*, "Hyperspectral pansharpening: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 27–46, Sep 2015.

[15] J. Zhou, D. Civco, and J. Silander, "A wavelet transform method to merge Landsat TM and SPOT panchromatic data," *Int. J. Remote Sens.*, vol. 19, no. 4, pp. 743–757, 1998.

[16] M. Selva, B. Aiazzi, F. Butera, L. Chiarantini, and S. Baronti, "Hypersharpening: A first approach on SIM-GA data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 3008–3024, Jun. 2015.

[17] G. Vivone *et al.*, "Pansharpening based on semiblind deconvolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 1997–2010, Apr. 2015.

[18] M. S. Karoui, Y. Deville, F. Z. Benhalouche, and I. Boukerch, "Hypersharpening by joint-criterion nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1660–1670, Mar. 2017.

[19] G. D. Robinson, H. N. Gross, and J. R. Schott, "Evaluation of two applications of spectral mixing models to image fusion," *Remote Sens. Environ.*, vol. 71, no. 3, pp. 272–281, Mar. 2000.

[20] B. Zhukov, D. Oertel, F. Lanzl, and G. Reinhackel, "Unmixing-based multisensor multiresolution image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1212–1226, May 1999.

[21] O. Berne, A. Helens, P. Pilleri, and C. Joblin, "Non-negative matrix factorization pansharpening of hyperspectral data: An application to mid-infrared astronomy," in *Proc. 2nd Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens.*, Reykjavik, Iceland, Jun. 2010, pp. 1–4.

[22] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, Feb. 2012.

[23] C. Lanaras, E. Baltsavias, and K. Schindler, "Hyperspectral superresolution by coupled spectral unmixing," in *Proc. IEEE Int. Conf. Comput. Vis.*, Santiago, Chile, Dec. 2015, pp. 3586–3594.

[24] K. Zhang, M. Wang, S. Y. Yang, Y. H. Xing, and R. Qu, "Fusion of panchromatic and multispectral images via coupled sparse non-negative matrix factorization," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 12, pp. 5740–5747, Dec. 2016.

[25] Q. Wei, J. B. Dias, N. Dobigeon, J. Y. Tourneret, M Chen, and S Godsill, "Multiband image fusion based on spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7236–7249, Dec. 2016.

[26] X. Zhang, J. Zhang, C. Li, C. Cheng, L. Jiao, and H. Zhou, "Hybrid unmixing based on adaptive region segmentation for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 3861–3875, Jul. 2018.

[27] R. Arablouei, "Fusion of multiple multiband images with complementary spatial and spectral resolutions," in *Proc. Int. Conf. Acoust., Speech Signal Process.*, Calgary, Canada, Apr. 2018, pp. 6478–6482.

[28] W. Fu, S. Li, L. Fang, X. Kang, and J. A. Benediktsson, "Hyperspectral image classification via shape-adaptive joint sparse representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 556–567, Feb. 2016.

[29] W. Dong *et al.*, "Hyperspectral image super-resolution via non-negative structured sparse representation," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2337–2352, May 2016.

[30] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[31] C. Han, H. Zhang, C. Gao, C. Jiang, N. Sang, and L. Zhang, "A remote sensing image fusion method based on the analysis sparse model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 1, pp. 439–453, Jan. 2016.

[32] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[33] H. Yin, S. Li, and L. Fang, "Simultaneous image fusion and superresolution using sparse representation," *Inf. Fusion*, vol. 14, no. 3, pp. 229–240, Jul. 2013.

[34] K. Zhang, M. Wang, and S. Yang, "Multispectral and hyperspectral image fusion based on group spectral embedding and low-rank factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1363–1371, Mar. 2017.

[35] M. Guo, H. Zhang, J. Li, L. Zhang, and H. Shen, "An online coupled dictionary learning approach for remote sensing image fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1284–1294, Apr. 2014.

[36] N. Akhtar, F. Shafait, and A. Mian, "Bayesian sparse representation for hyperspectral image super resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, Jun. 2015, pp. 3640–3641.

[37] N. Akhtar, F. Shafait, and A. Mian, "Sparse spatio-spectral representation for hyperspectral image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, Zurich, Switzerland, Sep. 2014, pp. 63–78.

[38] Q. Wei, N. Dobigeon, and J. Y. Tourneret, "Bayesian fusion of multi-band images," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1117–1127, Sep. 2015.

[39] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Fast fusion of multi-band images based on solving a Sylvester equation," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4109–4121, Nov. 2015.

[40] Q. Wei, J. B. Dias, N. Dobigeon, and J. Y. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3658–3668, Jul. 2015.

[41] G. Polatkan, M. Y. Zhou, L. Carin, D. Blei, and I. Daubechies, "A Bayesian nonparametric approach to image super resolution," *IEEE Trans. Pattern Anal. Mach. Learn.*, vol. 37, pp. 346–358, Feb. 2015.

[42] S. Ghosh and E. B. Sudderth, "Nonparametric learning for layered segmentation of natural images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, Jun. 2012, pp. 2272–2279.

[43] C. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[44] M. Afonso, J. Bioucas-Dias, and M. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, Mar. 2011.

[45] J. M. Bioucas-Dias and J. M. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.

[46] M. Sim˜oes, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3373–3388, Jun. 2015.

[47] R. C. Hardie, M. T. Eismann, and G. L. Wilson, "MAP estimation for hyperspectral image resolution enhancement using an auxiliary sensor," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1174–1184, Sep. 2004.

[48] M. Zhou, H. Chen, J. Paisley, L. Ren, G. Sapiro, and L. Carin, "Non-parametric Bayesian dictionary learning for sparse image representations 1," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 2295–2303.

[49] J. Paisley and L. Carin, "Nonparametric factor analysis with beta process priors," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 777–784.

[50] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. New York, NY, USA: Springer, 2004.

[51] J. Eckstein and D. P. Bertsekas, "On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, no. 1–3, pp. 293–318, Apr. 1992.

[52] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. Benediktsson, "SVM- and MRF-based method for accurate classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 736–740, Oct. 2010.

[53] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral-spatial classification of hyperspectral data using loopy belief propagation and active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 844–856, Feb. 2013.

[54] M. F. Baumgardner, L. L. Biehl, and D. A. Landgrebe, *220 Band AVIRIS Hyperspectral Image Data Set: June 12, 1992 Indian Pine Test Site 3, Version 1.0*. Sep. 2015. [Online]. Available: https://purr.purdue.edu/publications/1947/1

[55] M. Veganzones, M. Sim˜oes, G. Licciardi, N. Yokoya, J. Bioucas-Dias, and J. Chanussot, "Hyperspectral super-resolution of locally low rank images from complementary multisource data," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 274–288, Jan. 2016.
[56] R. Guerra, S. L´opez, and R. Sarmiento, "A computationally efficient algorithm for fusing multispectral and hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 5712–5728, Oct. 2016.
[57] E. M. Middleton *et al.*, "The earth observing one (EO-1) satellite mission: Over a decade in space," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 243–256, Apr. 2013.
[58] L. Wald, "Quality of high resolution synthesised images: Is there a simple criterion?" in *Proc. Int. Conf. Fusion Earth Data*, Nice, France, Jan. 2000, pp. 99–103.
[59] M. T. Eismann and R. C. Hardie, "Hyperspectral resolution enhancement using high-resolution multispectral imagery with arbitrary response functions," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 455–465, Mar. 2005.

Lichun Sui received the B.Eng. and M.Sc. degrees in photogrammetry from Wuhan University, China in 1983 and 1989, respectively, and the Ph.D. degree in photogrammetry and remote sensing from TU Berlin, Germany in 1998.

He is currently a Full Professor at the College of Geology Engineering and Geomatics, Changan University, China.

His main research interests comprise LiDAR data processing, photogrammetry and remote sensing application, remote sensing information extraction and application, image analysis and image processing, InSAR technology application, and 3D laser scanning technology in geological disaster monitoring.

Li Li received the B.Eng. degree in surveying and mapping engineering from Lanzhou Jiaotong University, Gansu, China, in 2010, and the M.Sc. degree in geodesy and the Ph.D. degree in photogrammetry and remote sensing, both from Changan University, Xian, China, in 2014 and 2018, respectively.

She is currently an engineer at China JIKAN Research Institute of Engineering Investigations and Design, Co., Ltd., Xian, China.

Her current research interests include image super-resolution and image fusion.

Jonathan Li (M'00–SM'11) received the Ph.D. degree in geomatics engineering from the University of Cape Town, South Africa, in 2000.

He is currently a Professor with both the Departments of Geography and Environmental Management as well as Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada.

His main research interests comprise LiDAR and SAR data processing, machine learning, and remote sensing applications. He has coauthored more than 400 publications, including over 180 refereed journal papers. He is currently serving as the Associate Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, and *Canadian Journal of Remote Sensing*.

Nan Chen received the B.Eng. degree in surveying and mapping engineering and the M.Sc. degree in photogrammetry and remote sensing from Chengdu University of Technology, China, in 2009 and 2012, respectively. She is currently working toward the Ph.D. degree in photogrammetry and remote sensing with Changan University, Xian, China.

Her current research interests include image super-resolution and image fusion.

Yongqing Jiao received the M.S. degree in cartography and geographic information engineering from Changan University, China, in 2013.

He is currently a Lecturer with both the Gansu Industry Polytechnic College and Gansu Engineering Research Center of Surveying and Mapping, China.

He is mainly engaged in surveying and mapping, and geographic information research and teaching.