

# Bayesian Hyperspectral and Multispectral Image Fusions via Double Matrix Factorization

Baihong Lin, Xiaoming Tao, *Member, IEEE*, Mai Xu, *Senior Member, IEEE*,  
Linhao Dong, *Member, IEEE*, and Jianhua Lu, *Fellow, IEEE*

**Abstract**—This paper focuses on fusing hyperspectral and multispectral images with an unknown arbitrary point spread function (PSF). Instead of obtaining the fused image based on the estimation of the PSF, a novel model is proposed without intervention of the PSF under Bayesian framework, in which the fused image is decomposed into double subspace-constrained matrix-factorization-based components and residuals. On the basis of the model, the fusion problem is cast as a minimum mean square error estimator of three factor matrices. Then, to approximate the posterior distribution of the unknowns efficiently, an estimation approach is developed based on variational Bayesian inference. Different from most previous works, the PSF is not required in the proposed model and is not pre-assumed to be spatially invariant. Hence, the proposed approach is not related to the estimation errors of the PSF and has potential computational benefits when extended to spatially variant imaging system. Moreover, model parameters in our approach are less dependent on the input data sets and most of them can be learned automatically without manual intervention. Exhaustive experiments on three data sets verify that our approach shows excellent performance and more robustness to the noise with acceptable computational complexity, compared with other state-of-the-art methods.

**Index Terms**—Double matrix factorization, image fusion, subspace-constrained image model, variational Bayesian inference.

## I. INTRODUCTION

SPECTRAL images play an important role in many computer vision tasks like classification and recognition, especially in remote sensing and medical imaging [1]. However, obtaining high spectral and spatial resolution versions of such images is not an easy task, since there always exists a tradeoff between the resolutions of spectra and space in the manufactured device. Moreover, constraints of communication and budget on the satellite aggravate the resolution due to the limited communication bandwidth and

Manuscript received September 5, 2016; revised December 5, 2016, March 13, 2017, and May 6, 2017; accepted May 30, 2017. Date of publication June 27, 2017; date of current version September 25, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 61622110, Grant 61471220, and Grant 91538107, and in part by the National Basic Research Project of China (973) under Grant 2013CB329006. (Corresponding author: Xiaoming Tao.)

B. Lin, X. Tao, L. Dong, and J. Lu are with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: linbaihong11@126.com; taoxm@tsinghua.edu.cn; dlh0601@hotmail.com; llh-dee@mail.tsinghua.edu.cn).

M. Xu is with the School of Electronics and Information Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191, China (e-mail: maixu@buaa.edu.cn).

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/TGRS.2017.2711640

energy cost, respectively [2], [3]. To meet the requirement in some applications, pan-sharpening, a postprocessing method, is developed, serving as a solution to yield the high-resolution multispectral image (MSI) by merging a low spatial resolution spectral image and a high spatial resolution panchromatic image [3]–[5]. Most recently, this technique has been extended to hyperspectral image (HSI) and MSI fusion after years of development [6]. As a result, a great challenge is launched that higher fusion quality and lower computational complexity are called up on the fusion methods, when facing problems of reconstructing images of tens or hundreds of spectral channels.

There have been many studies on the HSI fusion. As this issue is in fact an ill-posed inverse problem, a rational prior model on the high-resolution HSI is the corresponding bottleneck [6], [7]. Early researches mostly focus on projection and substitution [8], including principal component analysis [9] and wavelet decomposition [10]–[12]. They usually discuss the fusion problem under a low-dimensional spectral subspace. As obtaining the subspace is equivalent to dimensionality reduction, these methods are of low computational complexity. Unfortunately, they show unsatisfactory performance in fusion quality, as subspace regulation is a strong constraint for spectral image, which will complicate the coefficient modeling and easily cause spectral distortion [7]. Furthermore, it has been reported that this modeling is opposite to the physical imaging law [13].

To overcome the drawback above, various methods have been proposed [2]–[7], [13]–[24], among which the matrix factorization (MF) methods are the successful ones. This kind of methods factorizes a high-resolution image into double multiplied matrices [25], the spectral matrix, and the abundance matrix [19], which are based on the linear spectral mixture model used for spectral unmixing [20]. Then, by modeling the fusion issue as a biconvex optimization problem, the optimal values of two factor matrices can be obtained. Obviously, different priors on the factor matrices lead to different fusion methods, and so far as we have known, the widely adopted priors include sparse priors (e.g.,  $l_0$ -sparse-representation [21],  $l_1$ -Laplace priors [22], [23], and “spike-and-slab” priors [3]), low rank priors [16], [26],  $l_2$ -Gaussian priors [18], model-based priors [6], and so on.

However, three problems arise when most of the above methods are put into implementation. First, the relative point spread function (PSF) required in these methods, also called the blurring matrix, is usually defined based on the spatially invariant Gaussian hypothesis to accelerate the fusion

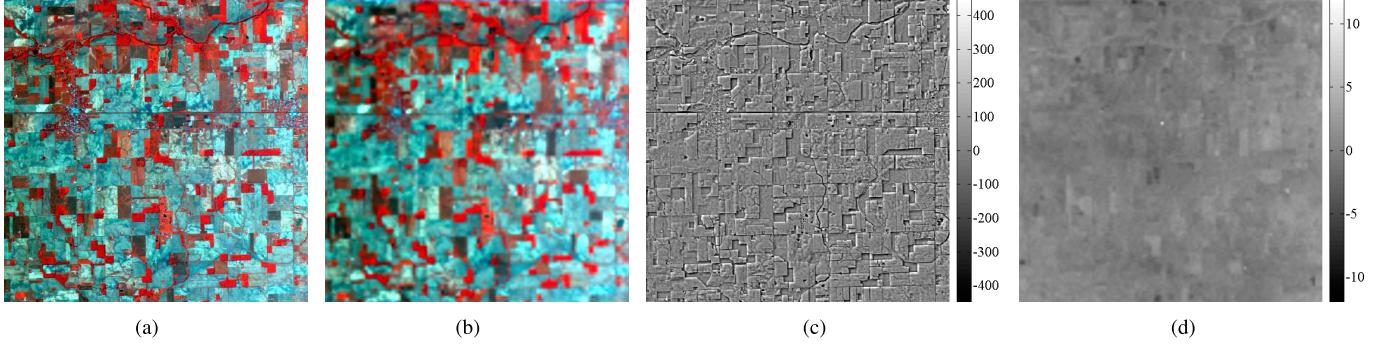


Fig. 1. (a) Proposed image model is decomposed into (b)–(d) three components. (a) and (b) False color. (a) High-resolution HSI,  $\mathbf{Z}$ , acquired by AVIRIS over Indian Pines with 220 spectral bands. (b) Degraded HSI,  $\mathbf{H}\mathbf{U}'\mathbf{W}$ . (c) Compensation component,  $\mathbf{H}\mathbf{U}'\mathbf{V}_{132}$ . (d) Residual component,  $\mathbf{R}_{132}$ .

methods [27]. But this hypothesis would probably aggravate the fusion performance for spatially varying imaging systems. Second, the relative PSF is hard to accurately measure in practice. Although previous works propose many methods to estimate the PSF [28]–[31], estimation errors vary with signal-to-noise ratio (SNR) and will be propagated to the fusion procedure, leading to deterioration of the fusion robustness to the noise. Furthermore, it consumes lots of memory to store the PSF for spatially varying imaging systems, since the PSFs at different pixel locations are different from each other. Third, results of these methods are dependent on the introduced regularization parameters that have to be set manually by experiments on relevant data sets. Once the parameters are not tuned carefully or the input images are not similar to that in the simulation, fusion qualities are poor despite that the results are optimal under the optimization framework.

In this paper, we discuss the fusion problem under an unknown arbitrary PSF. Instead of fusing images based on the estimated PSF, we develop a novel fusion model without intervention of the PSF. Specifically, the fused HSI is decomposed into a degraded image, a compensation component, and residuals (see Fig. 1), while the degraded image and the compensation component are modeled by double MF, sharing the same spectral feature matrix, restricted by an independently estimated spectral subspace. Then, we introduce priors for factor matrices, impose hyperpriors on hyperparameters, and cast the fusion problem as a minimum mean square error (MMSE) estimator under Bayesian framework. Finally, a variational inference approach is devised to efficiently obtain the fusion result. The main advantages of our framework can be summarized as follows.

- 1) A potentially spatially varying system PSF can be incorporated into the framework with no sacrifice in computational time, since no spatially invariant assumption is made on the PSF in our model.
- 2) As the PSF is not required in our model, the memory of the PSF can be saved, and there do not exist estimation errors of the PSF, let alone the influence of estimation errors on the fusion performance.
- 3) To avoid manually tuning hyperparameters, hyperpriors have been introduced. So all hyperparameters can be automatically learned to adapt to a wide range of SNRs during the inference process.

The remainder of this paper is organized as follows. Section II formulates the problem in mathematical forms and illustrates the proposed Bayesian model in detail. In Section III, we show our variational expectation-maximization (EM) approach and analyze the corresponding computational complexity. Experimental results and comparisons are given in Section IV and the conclusion is drawn in Section V.

## II. PROBLEM STATEMENT AND BAYESIAN MODELING

### A. Problem Description and Notations

Let  $\mathbf{X} \in \mathbb{R}^{L \times mn}$  denote a  $m$ -by- $n$  HSI with  $L$  channel bands and let  $\mathbf{Y} \in \mathbb{R}^{l \times MN}$  denote a high spatial resolution MSI with the size of  $M \times N \times l$ . Apparently, the image  $\mathbf{Y}$  features  $l$  bands and  $l < L$ ,  $n < N$ , and  $m < M$ . Then, the fused HSI is denoted by  $\mathbf{Z} \in \mathbb{R}^{L \times MN}$ . Many works have pointed out that the relationship among  $\mathbf{X}$ ,  $\mathbf{Y}$ , and  $\mathbf{Z}$  satisfies [28], [32]–[34]

$$\mathbf{X} = \mathbf{Z}\mathbf{G} + \mathbf{N}_X, \quad \text{and} \quad \mathbf{Y} = \mathbf{F}\mathbf{Z} + \mathbf{N}_Y \quad (1)$$

wherein  $\mathbf{F} \in \mathbb{R}^{l \times L}$  denotes the spectral response matrix, while  $\mathbf{G} \in \mathbb{R}^{MN \times mn}$  models the effect of both optic blurring and down-sampling.  $\mathbf{N}_X$  and  $\mathbf{N}_Y$  are additive perturbations caused by the measurements. In most researches, the PSF is assumed to be a spatially invariant kernel. In this case,  $\mathbf{G}$  can be factored into a block-circulant matrix with circulant blocks and a down-sampling matrix. Thus, the computation of matrix-vector multiplication can be accelerated using fast Fourier transformation [35], [36]. Unfortunately, the mentioned factorization of  $\mathbf{G}$  fails to work when the imaging system is spatially variant [31]. Moreover, in practice, it is hard to accurately obtain  $\mathbf{G}$  from the image data with noise, and estimation errors of  $\mathbf{G}$  can easily lead to poor robustness of fusion methods. As a result, this paper attempts to discuss the fusion problem, reconstructing the HSI,  $\mathbf{Z}$ , without intervention of  $\mathbf{G}$ . Since the problem is ill-posed, we choose the Bayesian methodology to analyze the problem due to its convenience and efficiency [18].

In Bayesian framework, the images  $\mathbf{X}$ ,  $\mathbf{Y}$ , and  $\mathbf{Z}$  are regarded as random variables. Let  $L(\mathbf{Z}, \hat{\mathbf{Z}})$  denote the loss function between the estimated image  $\hat{\mathbf{Z}}$  and the true image  $\mathbf{Z}$ . Then, the problem above turns to a Bayesian estimator that minimizes the loss expectation  $\mathbb{E}[L(\mathbf{Z}, \hat{\mathbf{Z}})|\mathbf{X}, \mathbf{Y}]$ . This paper focuses on the MMSE estimation, i.e.,  $L(\mathbf{Z}, \hat{\mathbf{Z}}) = \|\mathbf{Z} - \hat{\mathbf{Z}}\|_F^2$ ,

TABLE I  
NOTATIONS

Notation	Definition
$\text{tr}(\cdot)$	the trace of a matrix
$\otimes$	Kronecker product [37]
$\mathbf{I}_r$	a $r$ -by- $r$ identity matrix
$\mathbf{A} \in \mathbb{R}^{L \times N}$	a $L$ -by- $N$ real matrix
$\mathbf{A}_{\cdot j}$	the $j^{\text{th}}$ column of the matrix $\mathbf{A}$
$\mathbf{A}_{i \cdot}$	the $i^{\text{th}}$ row of the matrix $\mathbf{A}$
$\mathbf{A}'$	the transpose of the matrix $\mathbf{A}$
$\mathbf{A}^{-1}$	the inverse of the matrix $\mathbf{A}$
$\text{vec}[\mathbf{A}]$	vectorization operator, $[\mathbf{A}'_{\cdot 1}, \dots, \mathbf{A}'_{\cdot N}]'$
$\mathbb{E}(\cdot)$ or $(\cdot)$	the expectation of a random variable
$\dim[\mathbf{A}]$	the number of elements in $\mathbf{A}$ , $N \times L$
$\ \mathbf{A}\ _F$	the Frobenius matrix norm, $\sqrt{\text{tr}(\mathbf{A}\mathbf{A}')}$
$f(\cdot)$	the probabilistic density function (PDF)
$\mathcal{N}(\mu, \Sigma)$ or $\mathcal{N}(\mathbf{x} \mu, \Sigma)$	$\mathbf{x}$ has a multivariate Gaussian distribution with the mean of $\mu$ and the covariance matrix of $\Sigma$

and the estimation can be formulated as the following optimization problem:

$$\arg \min_{\widehat{\mathbf{Z}}} \int_{\mathbf{Z}} \|\widehat{\mathbf{Z}} - \mathbf{Z}\|_F^2 f(\mathbf{Z}|\mathbf{X}, \mathbf{Y}) \implies \widehat{\mathbf{Z}} = \mathbb{E}[\mathbf{Z}|\mathbf{X}, \mathbf{Y}].$$

Obviously, the joint distribution  $f(\mathbf{Z}, \mathbf{X}, \mathbf{Y})$  is the key to solve the above problem, and this distribution depends on the designed modeling shown in the following section. As there exist many mathematical operators in the following passage, for convenience, we list the notations of our paper in Table I.

### B. Proposed Subspace-Based Image Model

As has been pointed out, the spectral vector of each pixel in the HSI  $\mathbf{Z}$  lives in a much low-dimensional manifold. Previous works have proved that the manifold can be modeled by probabilistic MF (PMF) [33], [38], [39]. Let  $\mathbf{D} \in \mathbb{R}^{r \times L}$  and  $\mathbf{T} \in \mathbb{R}^{r \times MN}$  be latent spectral and texture feature matrices, with column vectors  $\mathbf{D}_{\cdot i}$  and  $\mathbf{T}_{\cdot j}$  representing spectral-specific and texture-specific latent feature vectors, respectively. Then,  $r$  represents the number of latent feature components. According to [38], [39]

$$\mathbf{Z} = \mathbf{D}'\mathbf{T} + \mathbf{N} \quad (2)$$

where  $\mathbf{N}$  denotes the residual out of the PMF model. However, in this paper, simply resorting to (1) and (2) cannot solve the fusion problem, since  $\mathbf{G}$  is unknown and is not expected to be estimated from the data as introduced in Section II-A. Furthermore, as  $\mathbf{X}$  and  $\mathbf{Y}$  are correlated by  $\mathbf{F}$  and  $\mathbf{G}$ , respectively in (1), posteriors of  $\mathbf{D}$  and  $\mathbf{T}$  obtained based on (1) and (2) are also related to  $\mathbf{G}$  and  $\mathbf{F}$ . These two posteriors tend to be extremely high-dimensional Gaussian distributions with big covariance matrices, which will result in unacceptable computational complexity in Bayesian inference.

To overcome the above problems, we introduce dimensionality reduction for  $\mathbf{D}$ , i.e.,  $\mathbf{D} = \mathbf{U}\mathbf{H}'$ ,  $\mathbf{U} \in \mathbb{R}^{r \times d}$ , and decompose  $\mathbf{T}$  into two texture feature matrices,  $\mathbf{W} \in \mathbb{R}^{r \times MN}$  and  $\mathbf{V} \in \mathbb{R}^{r \times MN}$ , as shown in the following equation:

$$\mathbf{Z} = \mathbf{H}\mathbf{U}'\mathbf{T} + \mathbf{R} = \mathbf{H}\mathbf{U}'(\mathbf{W} + \mathbf{V}) + \mathbf{R} \quad (3)$$

where  $\mathbf{H} \in \mathbb{R}^{L \times d}$  is a known spectral matrix obtained by truncated singular value decomposition (SVD) [40] or vertex component analysis (VCA) on the image  $\mathbf{X}$  [41] and  $\mathbf{R}$  is the residual out of framework. The proposed model has the following distinctive features.

First,  $\mathbf{U}$  is introduced to replace the latent spectral feature matrix  $\mathbf{D}$ . From (3), we can easily find that its element  $\mathbf{U}_{ij}$  represents how the  $j^{\text{th}}$  spectral basis  $\mathbf{H}_{\cdot j}$  makes contribution to the  $i^{\text{th}}$  latent texture feature component  $\mathbf{T}_{i \cdot}$ ,  $1 \leq i \leq r$ ,  $1 \leq j \leq d$ . Apparently, the row dimension of  $\mathbf{U}$  represents the number of latent feature components, while the column dimension of  $\mathbf{U}$  represents the dimension of spectral subspace  $\mathbf{H}$ . When  $r \leq d$ , the rank of  $\mathbf{D}$  mainly depends on  $r$ , and the proposed model is mathematically equivalent to the PMF model. According to [38], the larger  $r$ , the more powerful regression ability of the model. However, when  $r > d$ , latent spectral feature components  $\{\mathbf{D}'_1, \mathbf{D}'_2, \dots, \mathbf{D}'_r\}$  are linearly dependent due to the rank-deficiency of  $\mathbf{D}$  [42], and they are more like atoms of a spectral dictionary. Since no literature explores this special case, the corresponding performance will be further discussed based on experiments in Section IV-B.

Second,  $\mathbf{W}$  and  $\mathbf{V}$  are introduced to replace the latent texture feature matrix  $\mathbf{T}$ , and they are assumed to have different physical meanings:  $\mathbf{W}$  represents the latent texture feature in a degraded image while  $\mathbf{V}$  represents the compensation texture feature lost in a degraded image. This modeling in fact decomposes the high spatial resolution HSI into double matrix-factorization-based components,  $\mathbf{H}\mathbf{U}'\mathbf{W}$  and  $\mathbf{H}\mathbf{U}'\mathbf{V}$  (see Fig. 1):  $\mathbf{H}\mathbf{U}'\mathbf{W}$  is a degraded image mainly containing low spatial frequency components while  $\mathbf{H}\mathbf{U}'\mathbf{V}$  is a compensation matrix, which mainly contains high-frequency information (e.g., the edge compensation information of an image). Obviously, two different kinds of texture information are separated by this decomposition. In the following section, it can be found that this decomposition facilitates the modeling of sensor observations without intervention of  $\mathbf{G}$ .

### C. Sensor Observation Models via Matrix Factorizations

On the basis of the proposed image model, the prior distributions on the images  $\mathbf{X}$  and  $\mathbf{Y}$  are analyzed. According to (1), these two images can be regarded as projection observations of  $\mathbf{Z}$ . It is easy to find that  $\mathbf{X}$  mainly contains the component of  $\mathbf{H}\mathbf{U}'\mathbf{W}$ , while  $\mathbf{Y}$  contains the components of both  $\mathbf{H}\mathbf{U}'\mathbf{W}$  and  $\mathbf{H}\mathbf{U}'\mathbf{V}$ . Unfortunately, in this paper, the matrix  $\mathbf{G}$  is assumed unknown and its down-sampling effect widens the dimension gap between  $\mathbf{X}$  and  $\mathbf{H}\mathbf{U}'\mathbf{W}$ , leading to the difficulty of observation modeling. To overcome these issues,  $\mathbf{X}$  is upsampled. Let  $\widetilde{\mathbf{G}} \in \mathbb{R}^{MN \times MN}$  denote the effect of both  $\mathbf{G}$  and upsampling, and let  $\widetilde{\mathbf{N}}_X$  denote the sum of  $\mathbf{R}\widetilde{\mathbf{G}}$  and the upsampled perturbation for  $\mathbf{N}_X$ . Then, the resampled image  $\widetilde{\mathbf{X}} \in \mathbb{R}^{L \times MN}$  satisfies  $\widetilde{\mathbf{X}} = \mathbf{H}\mathbf{U}'\widetilde{\mathbf{G}} + \widetilde{\mathbf{N}}_X$ . As  $\widetilde{\mathbf{G}}$  can also be regarded as an unknown degradation function, according to Section II-B, the resampled image  $\widetilde{\mathbf{X}}$  can be directly taken as the degraded image ( $\mathbf{H}\mathbf{U}'\mathbf{W}$ ) added with the noise  $\widetilde{\mathbf{N}}_X$ .

Therefore, considering (1) and (3), the mathematical relationship of  $\widetilde{\mathbf{X}}$  and  $\mathbf{Y}$  is formulated as follows, where  $\widetilde{\mathbf{N}}_Y = \mathbf{N}_Y + \mathbf{F}\mathbf{R}$ :

$$\widetilde{\mathbf{X}} = \mathbf{H}\mathbf{U}'\mathbf{W} + \widetilde{\mathbf{N}}_X, \text{ and } \mathbf{Y} = \mathbf{F}\mathbf{H}\mathbf{U}'\mathbf{T} + \widetilde{\mathbf{N}}_Y. \quad (4)$$

Obviously, in (4),  $\tilde{\mathbf{N}}_X$  and  $\tilde{\mathbf{N}}_Y$  relate to each other for they all contain the residual  $\mathbf{R}$ . Although the residual is kept for analysis and is assumed to have a sparse prior in some works [39], [43], estimation of this variable will bring in a nonnegligible computational burden. Furthermore, according to Section II-B, the values of residual can be reduced by enhancing  $d$  and  $r$ , and it has few impact on the regression performance when  $d$  and  $r$  are large [38], [39]. Consequently, we neglect the residual and assume that  $\tilde{\mathbf{N}}_X$  and  $\tilde{\mathbf{N}}_Y$  are independent isotropic Gaussian noise, i.e.,  $\mathbf{N}_{Xij} \sim \mathcal{N}(0, \alpha_x^{-1})$  and  $\mathbf{N}_{Yij} \sim \mathcal{N}(0, \alpha_y^{-1})$ . Then, the distributions of sensor observations are found as follows:

$$\tilde{\mathbf{X}}|\mathbf{U}, \mathbf{W} \sim \prod_{j=1}^{MN} \mathcal{N}(\tilde{\mathbf{X}}_{\cdot j} | \mathbf{H}\mathbf{U}'\mathbf{W}_{\cdot j}, \alpha_x^{-1}\mathbf{I}_L) \quad (5)$$

$$\mathbf{Y}|\mathbf{U}, \mathbf{T} \sim \prod_{j=1}^{MN} \mathcal{N}(\mathbf{Y}_{\cdot j} | \mathbf{F}\mathbf{H}\mathbf{U}'\mathbf{T}_{\cdot j}, \alpha_y^{-1}\mathbf{I}_L). \quad (6)$$

#### D. Prior Distributions for Factor Matrices

Three factor matrices,  $\mathbf{U}$ ,  $\mathbf{V}$ , and  $\mathbf{W}$ , are unknown variables in the observation distributions (5) and (6). To make full use of their uncertain knowledge and avoid overfitting, prior distributions have to be introduced. We assume that these three matrices have isotropic Gaussian distributions, that is

$$\Theta \sim \prod_j \mathcal{N}(\Theta_{\cdot j} | \mathbf{0}, \alpha_\theta^{-1}\mathbf{I}_K), \quad \Theta \in \{\mathbf{U}, \mathbf{V}, \mathbf{W}\}. \quad (7)$$

These assumptions have been adopted in Bayesian RPCA [43] and PMF [38], [44]. Recent work even introduces them for HSI restoration [39]. As the size of  $\mathbf{U} \in \mathbb{R}^{r \times d}$  is hard to model, it will be set manually instead.

#### E. Hyperpriors on the Hyperparameters

Many hyperparameters are introduced in Section II-D. If they are all tuned manually, a complicate solution has to be given on the selection of parameters. Fortunately, hyperpriors can be introduced to learn the hyperparameters automatically in Bayesian framework. Let  $\Psi$  denote the set of all introduced hyperparameters. Then, Gamma distributions are employed for them, i.e.,  $\forall \alpha_x \in \Psi, \alpha_x \sim \Gamma(a_x, b_x)$

$$\Gamma(\alpha_x | a_x, b_x) \propto \alpha_x^{a_x-1} \exp[-\alpha_x b_x]. \quad (8)$$

The above parameters  $a_x$  and  $b_x$  are usually set to small values deterministically to avoid broad hyperpriors. In this paper, they are all set to  $10^{-6}$ .

Combining (5)–(8), we finally obtain the joint distribution as (9)

$$\begin{aligned} f(\tilde{\mathbf{X}}, \mathbf{Y}, \Omega, \Psi) &= f(\tilde{\mathbf{X}}|\mathbf{U}, \mathbf{W})f(\mathbf{Y}|\mathbf{U}, \mathbf{T}) \\ &\quad \times \prod_{\Theta \in \Omega} f(\Theta|\Psi) \prod_{\alpha \in \Psi} f(\alpha) \end{aligned} \quad (9)$$

$$\Omega = \{\mathbf{U}, \mathbf{V}, \mathbf{W}\}, \quad \Psi = \{\alpha_x, \alpha_y, \alpha_u, \alpha_v, \alpha_w\}.$$

### III. VARIATIONAL-EM-BASED FUSION APPROACH

According to Section II, as  $\mathbf{Z} \approx \mathbf{H}\mathbf{U}'(\mathbf{W} + \mathbf{V})$ , the calculation of  $\mathbb{E}[\mathbf{Z}|\mathbf{X}, \mathbf{Y}]$  is approximately equivalent to the calculation of  $\mathbb{E}[\Omega, \Psi|\tilde{\mathbf{X}}, \mathbf{Y}]$ . To obtain  $\mathbb{E}[\Omega, \Psi|\tilde{\mathbf{X}}, \mathbf{Y}]$ , the conditional posterior  $f(\Omega, \Psi|\tilde{\mathbf{X}}, \mathbf{Y})$  has to be given. Unfortunately, exact Bayesian inference for  $f(\Omega, \Psi|\tilde{\mathbf{X}}, \mathbf{Y})$  is not practical, since there exists an intractable integral,  $\int_{\Omega, \Psi} f(\tilde{\mathbf{X}}, \mathbf{Y}, \Omega, \Psi) d\Omega d\Psi$ , in the following inference process:

$$f(\Omega, \Psi|\tilde{\mathbf{X}}, \mathbf{Y}) = \frac{f(\tilde{\mathbf{X}}, \mathbf{Y}, \Omega, \Psi)}{\int_{\Omega, \Psi} f(\tilde{\mathbf{X}}, \mathbf{Y}, \Omega, \Psi) d\Omega d\Psi}. \quad (10)$$

Although the Markov chain Monte Carlo methods can theoretically estimate the exact value of the above expectation, its convergence is much slower and hard to guarantee [45]. Furthermore, it makes less sense to accurately obtain the value of  $\mathbb{E}[\Omega, \Psi|\tilde{\mathbf{X}}, \mathbf{Y}]$ , since all mathematical models in Section II are proposed statistically and they do not always work perfectly in the actual situation. Therefore, instead of exactly figuring out  $f(\Omega, \Psi|\tilde{\mathbf{X}}, \mathbf{Y})$ , we estimate the posterior approximately based on mean-field variational Bayes that is also called variational EM algorithm [45], [46].

The key idea of variational EM algorithm is to construct an approximate distribution, which has the minimal Kullback-Leibler divergence from the true posterior distribution. In the problem of this paper, let  $\hat{f}(\Omega, \Psi|\tilde{\mathbf{X}}, \mathbf{Y})$  denote the approximate posterior distribution. Then, the corresponding Kullback-Leibler divergence is written as

$$\mathcal{D}_{KL} = \int_{\Omega, \Psi} \hat{f}(\Omega, \Psi|\tilde{\mathbf{X}}, \mathbf{Y}) \log \frac{\hat{f}(\Omega, \Psi|\tilde{\mathbf{X}}, \mathbf{Y})}{f(\Omega, \Psi|\tilde{\mathbf{X}}, \mathbf{Y})} d\Omega d\Psi. \quad (11)$$

To solve the above minimal problem, the approximate distribution is usually assumed as the following form [45]:

$$\hat{f}(\Omega, \Psi|\tilde{\mathbf{X}}, \mathbf{Y}) = \prod_{\Theta \in \Omega} q(\Theta) \prod_{\psi \in \Psi} q(\psi) \quad (12)$$

where  $q(\cdot)$  is the approximate posterior of the corresponding random variable. This assumption is equivalent to the case that elements in  $\Omega \cup \Psi$  are independent of each other with known  $\tilde{\mathbf{X}}$  and  $\mathbf{Y}$ . Using the calculus of variations, the posterior  $q(\phi)$  of each  $\phi \in \Omega \cup \Psi$  can be formulated as follows:

$$\ln q(\phi) = \mathbb{E}[\ln f(\tilde{\mathbf{X}}, \mathbf{Y}, \Omega, \Psi) | \phi, \tilde{\mathbf{X}}, \mathbf{Y}] + \text{constant}. \quad (13)$$

Then, the posteriors of all random variables can be obtained by updating the above equation based on (9) iteratively. It is easy to find that this algorithm is more effective to avoid undesired local minima, since it approximates the posterior of each variable instead of finding peak point of the optimal function [47]. On the basis of (13), the approximate posterior of each variable is derived as follows.

#### A. Updating the Approximate Posterior of $\mathbf{U}$

From (13), the approximate posterior of  $\mathbf{U}$  is found as

$$\begin{aligned} \ln q(\mathbf{U}) &\propto -\frac{1}{2} \mathbb{E}[\alpha_y \|\mathbf{Y} - \mathbf{F}\mathbf{H}\mathbf{U}'\mathbf{T}\|_F^2 \\ &\quad + \alpha_x \|\tilde{\mathbf{X}} - \mathbf{H}\mathbf{U}'\mathbf{W}\|_F^2 + \alpha_u \|\mathbf{U}\|_F^2 | \tilde{\mathbf{X}}, \mathbf{Y}, \mathbf{U}, \Psi]. \end{aligned}$$

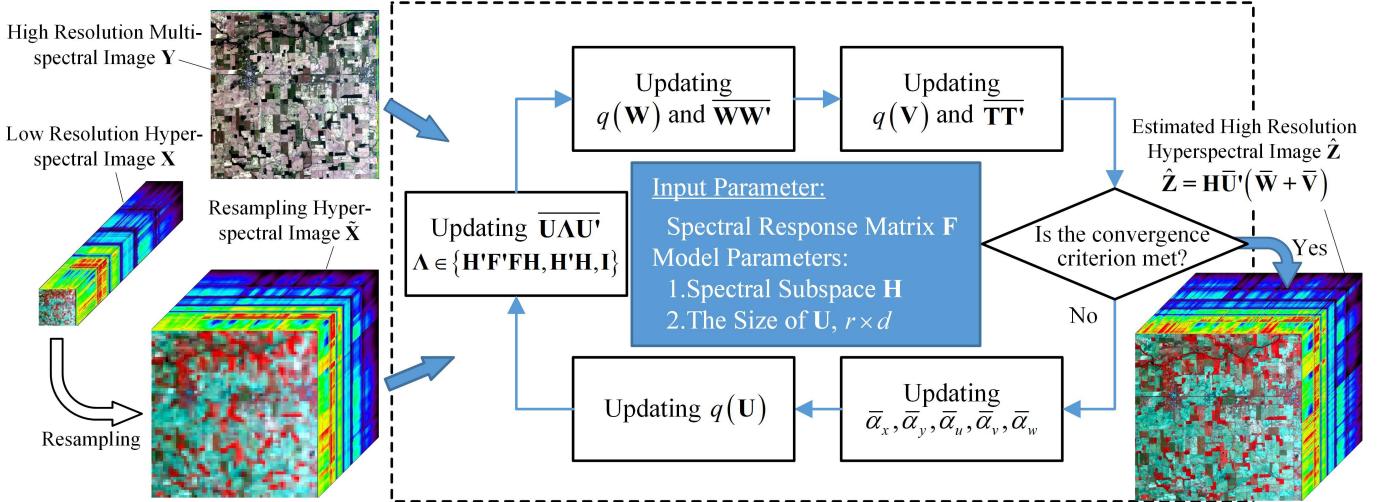


Fig. 2. Summary of the proposed variational-inference-based approach.

In the above equation, different columns of  $\mathbf{U}$  are correlated by  $\mathbf{H}$  and  $\mathbf{FH}$ , while different rows of  $\mathbf{U}$  are correlated by  $\mathbf{W}$  and  $\mathbf{T}$ . As the result,  $\mathbf{U}$  has to be vectorized so that its distribution can be easily modeled. Resorting to the vec-operator [48], the approximate posterior of  $\mathbf{U}$  is derived as

$$\begin{aligned} q(\mathbf{U}) &= q(\text{vec}[\mathbf{U}]) \sim \mathcal{N}(\text{vec}[\bar{\mathbf{U}}], \Sigma_u^*) \\ \text{vec}[\bar{\mathbf{U}}] &= \Sigma_u^* \text{vec}[\bar{\alpha}_x \bar{\mathbf{W}} \bar{\mathbf{X}}' \mathbf{H} + \bar{\alpha}_y \bar{\mathbf{T}} \mathbf{Y}' \mathbf{F}\mathbf{H}] \\ \Sigma_u^* &= (\bar{\alpha}_x \mathbf{H}' \mathbf{H} \otimes \bar{\mathbf{W}} \mathbf{W}' + \bar{\alpha}_y \mathbf{H}' \mathbf{F}' \mathbf{F}\mathbf{H} \otimes \bar{\mathbf{T}} \mathbf{T}' + \bar{\alpha}_u \mathbf{I}_{dr})^{-1}. \end{aligned} \quad (14)$$

$\bar{\mathbf{U}}\Lambda\bar{\mathbf{U}}'$  is required for other approximate posteriors, wherein  $\Lambda \in \mathbb{R}^{d \times d}$  is a known matrix and  $\Lambda = \mathbf{H}'\mathbf{F}'\mathbf{F}\mathbf{H}$  or  $\mathbf{H}'\mathbf{H}$  or  $\mathbf{I}$ . But it cannot be derived by easily substituting the expectation of each variable due to the correlation among elements in the matrix-variable. To calculate  $\bar{\mathbf{U}}\Lambda\bar{\mathbf{U}}'$ , the covariance  $\Sigma_u^* \in \mathbb{R}^{dr \times dr}$  has to be partitioned into  $d^2 r \times r$  sub-matrices  $[\Sigma_u^*]_{r \times r}^{ij}$ , which can be written as

$$\Sigma_u^* = \begin{bmatrix} [\Sigma_u^*]_{r \times r}^{11} & \cdots & [\Sigma_u^*]_{r \times r}^{1d} \\ \vdots & \ddots & \vdots \\ [\Sigma_u^*]_{r \times r}^{d1} & \cdots & [\Sigma_u^*]_{r \times r}^{dd} \end{bmatrix}.$$

Then, using the above partitioned matrix,  $\bar{\mathbf{U}}\Lambda\bar{\mathbf{U}}'$  is found as follows:

$$\bar{\mathbf{U}}\Lambda\bar{\mathbf{U}}' = \bar{\mathbf{U}}\Lambda\bar{\mathbf{U}}' + \sum_{i=1}^d \sum_{j=1}^d \Lambda_{ij} [\Sigma_u^*]_{r \times r}^{ij}. \quad (15)$$

### B. Updating the Approximate Posteriors of $\mathbf{W}$ and $\mathbf{V}$

As  $\mathbf{T} = \mathbf{W} + \mathbf{V}$ ,  $\mathbf{W}$  and  $\mathbf{V}$  feature similar locations in (5) and (6). According to (13), their posteriors have similar forms

and can be represented by

$$\begin{aligned} \ln q(\mathbf{W}) &\propto -\frac{1}{2} \mathbb{E}[\alpha_y \|\mathbf{Y} - \mathbf{F}\mathbf{H}\mathbf{U}'(\mathbf{W} + \mathbf{V})\|_F^2 \\ &\quad + \alpha_x \|\bar{\mathbf{X}} - \mathbf{H}\mathbf{U}'\mathbf{W}\|_F^2 + \alpha_w \|\mathbf{W}\|_F^2 \mid \bar{\mathbf{X}}, \mathbf{Y}, \mathbf{W}, \Psi] \\ \ln q(\mathbf{V}) &\propto -\frac{1}{2} \mathbb{E}[\alpha_y \|\mathbf{Y} - \mathbf{F}\mathbf{H}\mathbf{U}'(\mathbf{W} + \mathbf{V})\|_F^2 \\ &\quad + \alpha_v \|\mathbf{V}\|_F^2 \mid \bar{\mathbf{X}}, \mathbf{Y}, \mathbf{V}, \Psi]. \end{aligned}$$

The above equations show that different columns in  $\mathbf{W}$  and  $\mathbf{V}$  are independent of each other, and elements in a column are drawn from the corresponding Gaussian distributions found as follows:

$$q(\mathbf{W}_{\cdot j}) \sim \mathcal{N}(\bar{\mathbf{W}}_{\cdot j}, \Sigma_w^*) \quad (16)$$

$$\begin{aligned} \Sigma_w^* &= (\bar{\alpha}_x \bar{\mathbf{U}} \mathbf{H}' \mathbf{H} \bar{\mathbf{U}}' + \bar{\alpha}_y \bar{\mathbf{U}} \mathbf{H}' \mathbf{F}' \mathbf{F}\mathbf{H} \bar{\mathbf{U}}' + \bar{\alpha}_w \mathbf{I}_r)^{-1} \\ \bar{\mathbf{W}} &= \Sigma_w^* [\bar{\mathbf{U}} \mathbf{H}' (\bar{\alpha}_x \bar{\mathbf{X}} + \bar{\alpha}_y \mathbf{F}' \mathbf{Y}) + \bar{\alpha}_y \bar{\mathbf{U}} \mathbf{H}' \mathbf{F}' \mathbf{F}\mathbf{H} \bar{\mathbf{U}}' \cdot \bar{\mathbf{V}}] \\ q(\mathbf{V}_{\cdot j}) &\sim \mathcal{N}(\bar{\mathbf{V}}_{\cdot j}, \Sigma_v^*) \quad (17) \end{aligned}$$

$$\begin{aligned} \Sigma_v^* &= (\bar{\alpha}_y \bar{\mathbf{U}} \mathbf{H}' \mathbf{F}' \mathbf{F}\mathbf{H} \bar{\mathbf{U}}' + \bar{\alpha}_v \mathbf{I}_r)^{-1} \\ \bar{\mathbf{V}} &= \bar{\alpha}_y \Sigma_w^* [\bar{\mathbf{U}} \mathbf{H}' \mathbf{F}' \mathbf{Y} + \bar{\mathbf{U}} \mathbf{H}' \mathbf{F}' \mathbf{F}\mathbf{H} \bar{\mathbf{U}}' \cdot \bar{\mathbf{W}}]. \end{aligned}$$

From (16) and (17), it can be found that elements in each column of  $\mathbf{W}$  (or  $\mathbf{V}$ ) are not independent to each other. Thus, the calculation of  $\bar{\mathbf{W}}\bar{\mathbf{W}}'$  cannot be obtained by simply multiplying  $\bar{\mathbf{W}}$  and  $\bar{\mathbf{W}}'$  together. So does the calculation of  $\bar{\mathbf{V}}\bar{\mathbf{V}}'$ . Considering the covariance  $\Sigma_w^*$  and  $\Sigma_v^*$ , the values of  $\bar{\mathbf{W}}\bar{\mathbf{W}}'$  and  $\bar{\mathbf{V}}\bar{\mathbf{V}}'$  in (14) should be calculated as follows, in which the dot operator denotes matrix multiplication:

$$\bar{\mathbf{W}}\bar{\mathbf{W}}' = \bar{\mathbf{W}} \cdot \bar{\mathbf{W}}' + MN \Sigma_w^* \quad (18)$$

$$\bar{\mathbf{V}}\bar{\mathbf{V}}' = (\bar{\mathbf{W}} + \bar{\mathbf{V}})(\bar{\mathbf{W}} + \bar{\mathbf{V}})' + MN(\Sigma_w^* + \Sigma_v^*). \quad (19)$$

### C. Updating the Approximate Posteriors of Hyperparameters

As Gamma distribution is conjugate to Gaussian distribution, the approximate posteriors of all hyperparameters are still found as Gamma distributions [45]. From (8) and (13), means

of hyperparameters are calculated as

$$\bar{\alpha}_x = \frac{10^{-6} + MNL/2}{10^{-6} + \mathbb{E}\|\tilde{\mathbf{X}} - \mathbf{H}\mathbf{U}'\mathbf{W}\|_F^2/2} \quad (20)$$

$$\bar{\alpha}_y = \frac{10^{-6} + MNl/2}{10^{-6} + \mathbb{E}\|\mathbf{Y} - \mathbf{F}\mathbf{H}\mathbf{U}'\mathbf{T}\|_F^2/2} \quad (21)$$

$$\bar{\alpha}_\theta = \frac{10^{-6} + \dim[\Theta]/2}{10^{-6} + \text{tr}[\Theta\Theta']/2}, \quad \Theta \in \Omega, \quad \alpha_\theta \in \{\alpha_u, \alpha_v, \alpha_w\}. \quad (22)$$

The steps of the proposed variational approach are summarized in Fig. 2. The high spatial resolution HSI is finally estimated based on the means of all factor matrices as  $\hat{\mathbf{Z}} = \mathbf{H}\mathbf{U}'(\overline{\mathbf{W}} + \overline{\mathbf{V}})$ .

#### D. Analysis of Computational Complexity

The approximations of the above posterior distributions have different orders of computational complexity: Updating  $q(\mathbf{U})$  has the order of complexity  $\mathcal{O}(rdMN)$ , while updating  $q(\mathbf{W})$ ,  $q(\mathbf{V})$ , and posteriors of all hyperparameters has the same order of complexity  $\mathcal{O}(LMN \max(d, r))$ . Let  $\Delta$  denote the iteration number, and  $\max(d, r) < L$ . Then, considering all above orders of complexity, the proposed approach has the order of complexity  $\mathcal{O}(LMN\Delta \max(d, r))$ .

## IV. EXPERIMENTAL RESULTS AND ANALYSIS

### A. Experiment Scenario and Evaluation

In order to use a reference image for quality assessment, a semi-synthetic simulation method based on Wald's protocol [49] is used for experiments [32]. In the simulation, a real-life HSI is taken as the ground truth. Low spatial resolution HSI is generated by downsampling the blurred version of the real-life HSI, while the high spatial resolution MSI is obtained by simulating the spectral response on the ground-truth image.

In our experiment, to validate the proposed method, three real-life data sets with different spectral channel bands are adopted, including the Balloons, the Pavia University (PaviaU), and the Moffett data sets. The first data set, the Balloons, is a  $512 \times 512 \times 31$  HSI with 31 spectral channel bands, ranging from 400 to 700 nm at 10-nm intervals in wavelength. It is acquired by a generalized assorted pixel camera [50]. As a classic static scene of the CAVE database [50], this data set has been widely used for evaluation of HS and RGB image fusion in many similar works [19], [33], [34], [51]. We generate the corresponding RGB images based on the Nikon D700 camera's spectral response.<sup>1</sup> Then, a low spatial resolution  $64 \times 64 \times 31$  HSI is created by adopting a  $16 \times 16$  Gaussian kernel with the variance of 3.40.

The other two data sets, the PaviaU<sup>2</sup> and the Moffett data sets, are obtained by the reflective optics system imaging spectrometer (ROSIS)<sup>3</sup> and the NASA's Airborne Visible

Infrared Imaging Spectrometer (AVIRIS),<sup>4</sup> respectively. These two data sets are popular in remote sensing image processing. The former has 115 spectral channel bands ranging from 430 to 860 nm, while the latter has 224 spectral channel bands ranging 400–2500 nm with 10-nm intervals. To ensure the image quality, low-quality spectral bands caused by the water absorptions have to be removed, which reduces the spectral bands of these two images to 93 and 188. For simplicity, we set the spatial size of these two images to  $240 \times 240$  and generate the corresponding HSIs with the  $48 \times 48$  spatial size using a  $9 \times 9$  Gaussian kernel with a variance of 2.12. IKONOS-like spectral response is adopted to generate the MSI for the PaviaU image while six bands of the Moffett image corresponding to the wavelengths 480, 560, 660, 830, 1650, and 2220 nm are directly chosen to simulate the MSI acquired by the USGS/NASA Landsat 7 satellite.

To simulate the perturbations for three data sets, white Gaussian noise is added to generated images according to (1), and the SNRs of HS and MS (or RGB) images are denoted by  $\text{SNR}_h$  and  $\text{SNR}_m$ , respectively. Then, to evaluate the difference between the fused image  $\hat{\mathbf{Z}}$  and the ground truth  $\mathbf{Z}$ , four quality assessment metrics are introduced as below.

1) *Root Mean Squared Error*: The RMSE has been widely used in image processing. This quality measurement can be computed with the following equation:

$$\text{RMSE}(\mathbf{Z}, \hat{\mathbf{Z}}) = \frac{\|\mathbf{Z} - \hat{\mathbf{Z}}\|_F}{\sqrt{MNL}}. \quad (23)$$

2) *Spectral Angle Mapper*: The SAM measures the spectral distortion between the reference image and the fused image. If the fused image is ideal, the value of SAM is zero. Otherwise, it is obtained by averaging all spectral distortion angle of all pixels in the image, as shown in the following equation:

$$\text{SAM}(\mathbf{Z}, \hat{\mathbf{Z}}) = \frac{1}{MN} \sum_j \arccos \left( \frac{\hat{\mathbf{Z}}_{j,j}^T \mathbf{Z}_{j,j}}{\|\mathbf{Z}_{j,j}\|_2 \|\hat{\mathbf{Z}}_{j,j}\|_2} \right). \quad (24)$$

3) *Erreur Relative Globale Adimensionnelle De Synthèse*: The ERGAS is proposed to overcome the sensitivity to the changes from numerical counts to radiances. This metric is more robust in terms of calibration and changes of units. It is defined as [52]

$$\text{ERGAS}(\mathbf{Z}, \hat{\mathbf{Z}}) = \frac{100}{s} \sqrt{\frac{1}{L} \sum_{i=1}^L \left( \frac{\text{RMSE}(i)}{\mu_i} \right)^2}. \quad (25)$$

In the above definition,  $\text{RMSE}(i)$  and  $\mu_i$  denote the RMSE and the mean pixel value in the  $i$ th band of HSI, respectively.  $s$  is the spatial scale between the HSI and the MSI.

4) *Mean of Structural Similarity Index*: The structural similarity (SSIM) index measures the difference of two images patches,  $\mathbf{x}$  and  $\hat{\mathbf{x}}$ , with (26) [53]

$$\text{SSIM}(\mathbf{x}, \hat{\mathbf{x}}) = \frac{(2\mu_x\mu_{\hat{x}} + c_1)(2\sigma_x\sigma_{\hat{x}} + c_2)}{(\mu_x^2 + \mu_{\hat{x}}^2 + c_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + c_2)} \quad (26)$$

<sup>1</sup>Available at [http://www.maxmax.com/spectral\\_response.htm](http://www.maxmax.com/spectral_response.htm)

<sup>2</sup>The Pavia University image can be found at [http://www.ahu.eus/ccwintco/index.php?title=Hyperspectral\\_Remote\\_Sensing\\_Scenes](http://www.ahu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes).

<sup>3</sup>More information about ROSIS can be found at the following website: <http://messtec.dlr.de/en/technology/dlr-remote-sensing-technology-institute/hyperspectral-systems-airborne-rosis-hypspx/index.php>.

<sup>4</sup>More information about AVIRIS can be found at <http://aviris.jpl.nasa.gov/>.

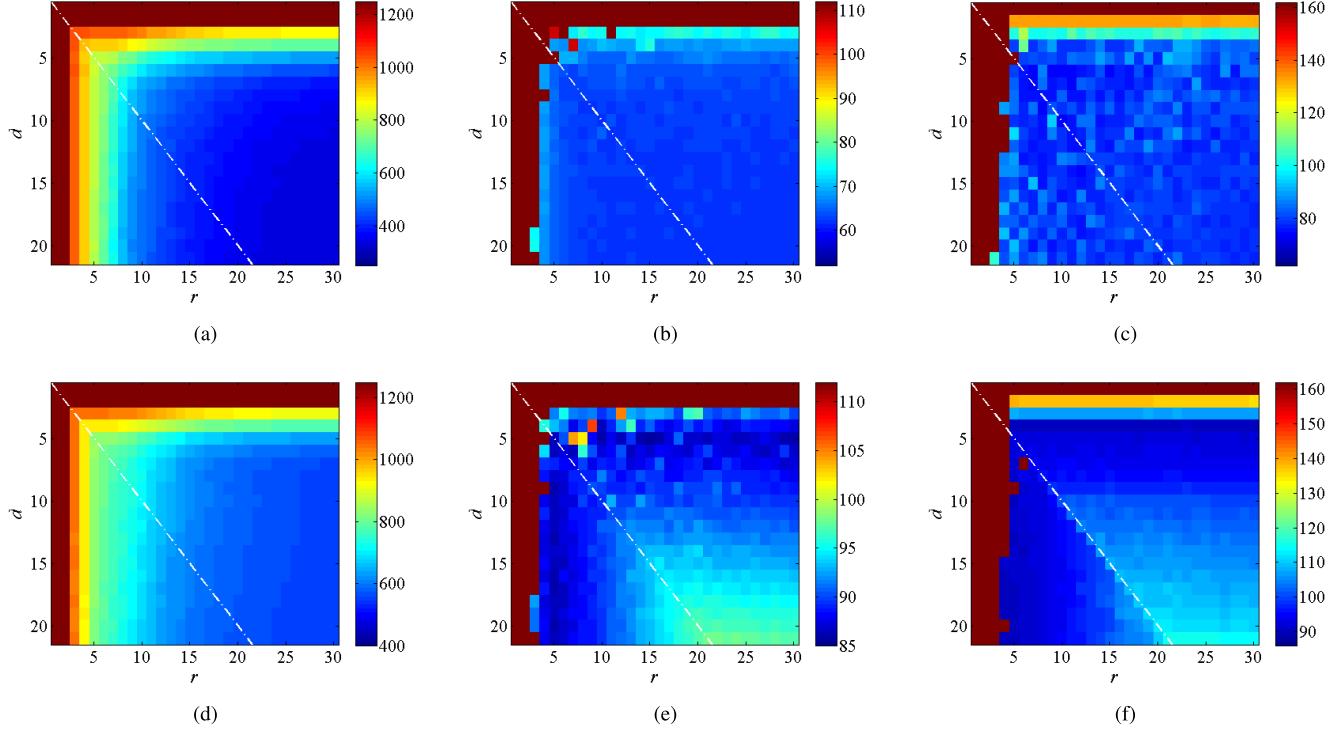


Fig. 3. RMSE values for different values of  $d$  and  $r$ .  $\tilde{\mathbf{X}}$  is obtained by bicubic interpolation. (a)–(c)  $\text{SNR}_h = \text{SNR}_m = \infty$ . (d)–(f)  $\text{SNR}_h = 20$  dB,  $\text{SNR}_m = 30$  dB. The white line represents  $d = r$ . (a) and (d) Balloons data sets results. (b) and (e) PaviaU data sets results. (c) and (f) Moffett data sets results.

where  $\mu_{\mathbf{x}}$  and  $\mu_{\hat{\mathbf{x}}}$  are the respective mean pixel values of the patch  $\mathbf{x}$  and  $\hat{\mathbf{x}}$ , respectively,  $\sigma_{\mathbf{x}}$  and  $\sigma_{\hat{\mathbf{x}}}$  are their corresponding variance, and  $\sigma_{\mathbf{x}\hat{\mathbf{x}}}$  is the covariance between two patches.  $c_1$  and  $c_2$  are introduced parameters that satisfies  $c_i = \sqrt{k_i}\eta$ ,  $i = 1, 2$ .  $\eta$  denotes dynamic range of the images. In this paper,  $k_1 = 0.01$ ,  $k_2 = 0.03$ . As all the metrics are measured in the range of 16-b images,  $\eta = 2^{16} - 1$ . To evaluate two images, a sliding window of  $32 \times 32$  pixels is used to segment pairs of patches at the same position in two images, which moves pixel by pixel horizontally and vertically. Then, the mean of SSIM index is obtained by averaging the SSIMs of all patches' pairs.

#### B. Implementation Details

The proposed algorithm is run on MATLAB R2013a with an Intel Core i7 CPU at 3.6 GHz and 16 GB RAM. Three problems arise in implementation. First, there are a lot of intermediate variables that are involved in the iteration loop. It is intractable to initiate them. Second, two model parameters are required for the proposed methods, including the spectral subspace  $\mathbf{H}$  and the size of  $\mathbf{U}$ . But the tuning of their values remains open to question. Third,  $\tilde{\mathbf{X}}$  is an upsampled image. The choice of its upsampling approach for different blur kernels still requires further exploration. This section mainly discusses the solutions to these three problems.

*1) Initialization:* We propose a Gibbs-sampling-based method to initiate the intermediate variables. In the initialization, samples of each variable are generated one by one from its corresponding conditional distribution, which

can be obtained based on (9). After 10 sampling rounds, the conditional distribution of each variable is taken as the initial approximate posterior for the proposed approach. This initialization method is equivalent to running a few rounds of Gibbs-sampling. Obviously, according to the Gibbs-sampling method [54], generated samples can be regarded to be drawn from the joint distribution when sampling rounds tend to infinity. Although the sampling rounds in our initialization are limited, the samples are more likely to be near the peak of the joint probability density function, enough for the initialization to accelerate the convergence.

*2) Selection of Model Parameters:* On the basis of the initialization, selections of the input parameters are analyzed. As has been reported, truncated SVD on the low-resolution HSI corresponds to the maximum likelihood estimation of spectral subspace if the image noise is i.i.d. or zero [28]. Otherwise, the calculation of spectral subspace will be complex [55]. In Section II-C, we have assumed that the perturbations are band-independent isotropic Gaussian noises. As a result, we utilize truncated SVD on low spatial resolution HSI,  $\mathbf{X}$ , to obtain the spectral subspace  $\mathbf{H}$ .

Then, to select the size of  $\mathbf{U}$ , two groups of experiments on the fusion of HSI and MSI are designed with various values of  $d$  and  $r$ . In the first group of experiments, all input images are generated under two different SNR conditions based on the scenarios in Section IV-A, and the spectral response matrix is assumed to be accurately known. The corresponding results are shown in Figs. 3 and 4.

Fig. 3 shows the relationship between the fusion quality and the size of  $\mathbf{U}$ . From Fig. 3(a)–(c), we can find that the

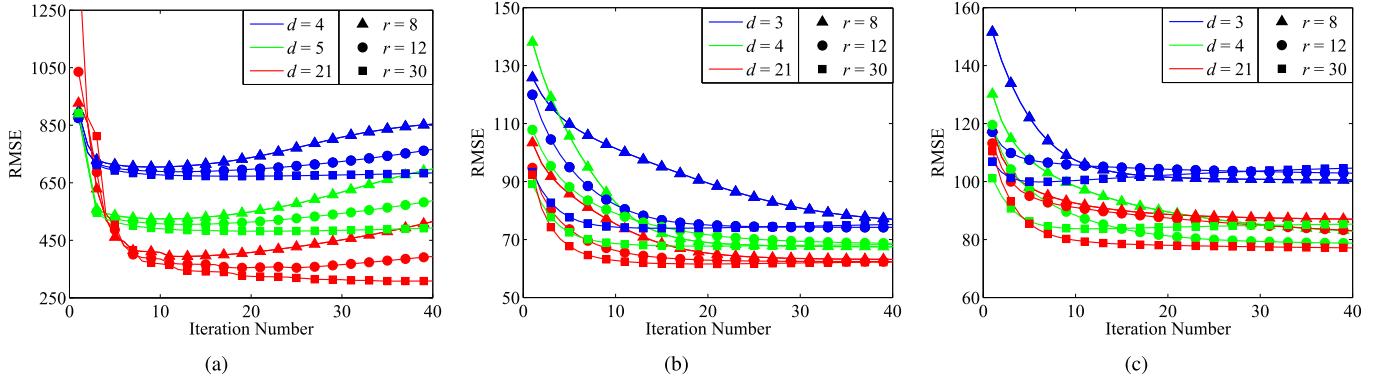


Fig. 4. RMSE values for different iteration numbers.  $\text{SNR}_h = \text{SNR}_m = \infty$ .  $\tilde{\mathbf{X}}$  is obtained by bicubic interpolation. (a)–(c) Results of the Balloons data set, the PaviaU data set, and the Moffett data set, respectively.

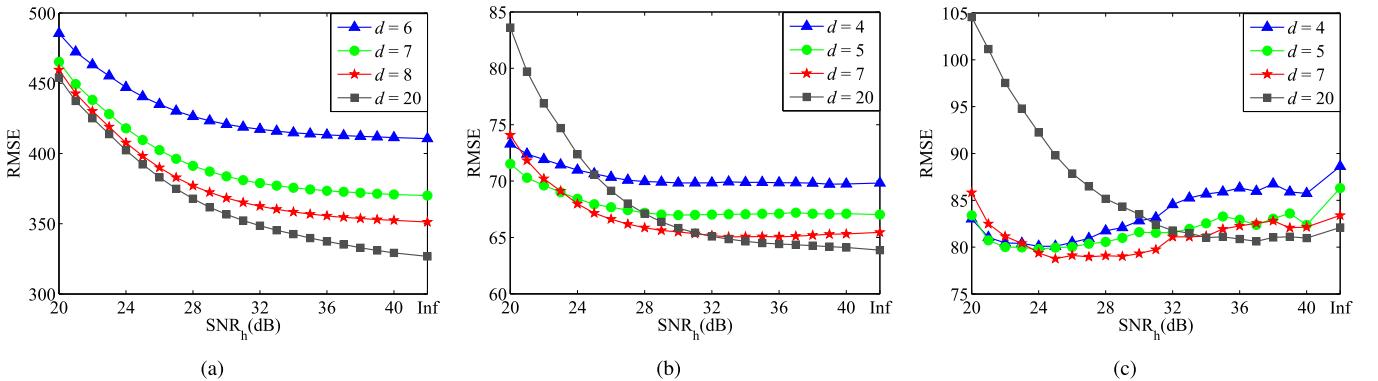


Fig. 5. RMSE values for different values of  $\text{SNR}_h$ .  $\text{SNR}_m = 40$  dB.  $r = 30$ .  $\tilde{\mathbf{X}}$  is obtained by bicubic interpolation. (a)–(c) Results of the Balloons data set, the PaviaU data set, and the Moffett data set, respectively.

values of RMSE decrease significantly with the increase in  $r$  and  $d$ , even when  $r > d$ ; the fusion qualities tend to be stable when  $d \geq 6$  and  $r \geq 7$ . These results validate the analysis on  $\mathbf{U}$  in Section II-B. However, the above trend changes when the input images are obtained at low SNRs ( $\text{SNR}_h = 20$  dB,  $\text{SNR}_m = 30$  dB). In Fig. 3(d), the values of RMSE tend to be better when  $d > 6$ ,  $r > 12$ , and  $r > d$ . These results imply that the proposed approach would have a good fusion quality if  $\mathbf{U}$  is a thin matrix ( $r > d$ ). However, in Fig. 3(e) and (f), the values of RMSE tend to be worse with the increase in  $d$  and  $r$  when  $d > 11$  and  $r > 10$ . This phenomenon is mainly caused by the overfitting of our model, which indicates that the values of  $d$  and  $r$  cannot be too large at the same time if the images are measured at low SNRs. Considering the data trends in Fig. 3, it would be better to let  $6 \leq d < 11$  and  $r > 12$ .

Fig. 4 shows the convergence curves with different sizes of  $\mathbf{U}$ . Obviously, in Fig. 4(a), our approach does not converge within 40 iterations for some combinations of  $r$  and  $d$ . But the curves in the figure tend to be better with the increase in  $r$ , especially when  $r = 30$ . As for Fig. 4(b) and (c), it is easy to find that the convergence speeds of the proposed approach tend to increase with the value in  $r$ , while  $d$  has a significant impact on the lower bound of convergence RMSE value. Furthermore, we can find that the fusion performance and the convergence speeds tend to be better when  $r > d$ . Therefore, to guarantee the convergence and improve the corresponding speed, it would be better to let  $r \geq 30$  and set  $\mathbf{U}$  as a thin

matrix ( $r > d$ ). Besides, in Fig. 4, the initial RMSE values of three data sets are close to their corresponding convergence values, which implies that the proposed initialization method is effective for these three data sets.

In the second group of experiments, HS images are generated under different  $\text{SNR}_h$  values based on the scenarios in Section IV-A. We still assume that the spectral response matrix is accurately known for the proposed methods. Inspired by Fig. 4, we let  $r = 30$ , and set the iteration of our approach to 30. The results are shown in Fig. 5, in which RMSE curves are smoothed with a 7-span moving average filter. It is easy to find that Fig. 5(a)–(c) has two different trends. In Fig. 5(a), the values of RMSE tend to be better with the increase in  $d$  and tend to be relatively stable when  $d \geq 8$ ; in Fig. 5(b) and (c), the values of RMSE show a reverse trend when  $d \geq 7$  and  $\text{SNR}_h < 24$  dB. These trends are similar to those in Fig. 3(d)–(f), and indicate that  $d$  should be set within the range of 7 to 20 and should be close to 7, when  $r = 30$ , and the images are obtained at low  $\text{SNR}_h$ .

Considering the balance among the fusion quality, the convergence speed, and the noise-resistance, it would be better to let  $7 \leq d < 11$ ,  $r \geq 30$ . In the following experiments, the size of  $\mathbf{U}$  is set to  $30 \times 10$  for all data sets, i.e.,  $d = 10$  and  $r = 30$ . As the proposed approach has already shown good performance with about 20 iterations, a stopping criterion is proposed that our approach is run for 20 iterations in every case.

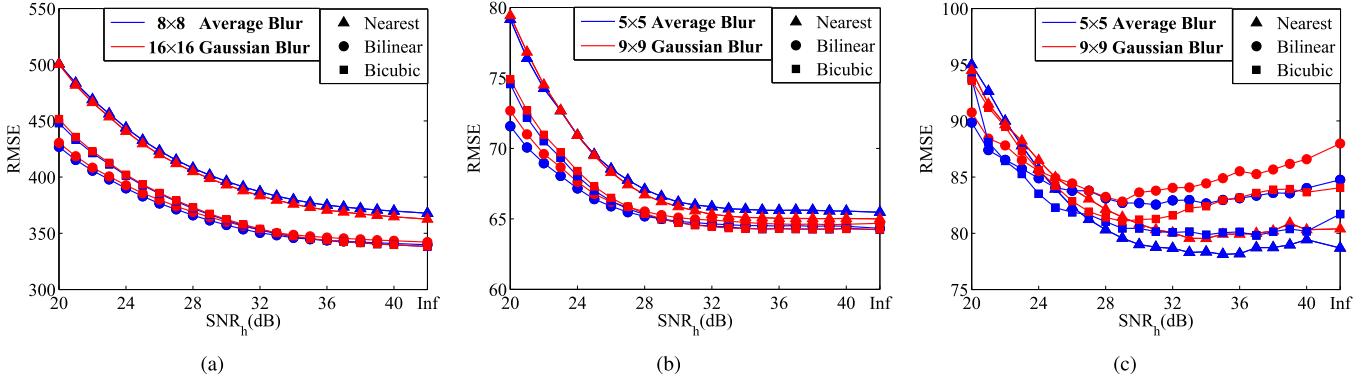


Fig. 6. RMSE values of the proposed method for different values of  $\text{SNR}_h$ , blur kernels, and the upsampling approaches.  $\text{SNR}_m = 40 \text{ dB}$ .  $d = 10$ ,  $r = 30$ . Three markers represent nearest neighbor, bilinear, and bicubic interpolation for  $\tilde{\mathbf{X}}$ , respectively. Standard deviations of the Gaussian blur kernels in three above subfigures are 3.40, 2.12, and 2.12, respectively. (a)–(c) Results of the Balloons data set, the PaviaU data set, and the Moffett data set, respectively.

3) *Selection of the Upsampling Method for  $\tilde{\mathbf{X}}$* : To explore the influence of upsampling methods on fusion performance, another group of experiments is designed. In these experiments, the spectral response matrix is assumed to be accurately known. Two kinds of blur kernels are adopted to generate the low resolution HSI, including the average blur kernel and the Gaussian blur kernel introduced in Section IV-A. These two kinds of blur kernels have been widely used in many similar works [2], [3], [22], [23]. Their curves of fusion quality are shown with different upsampling approaches for  $\tilde{\mathbf{X}}$  in Fig. 6. From the results, we can find that there only exists a little difference between the RMSE values of average and Gaussian blur kernels. Moreover, the proposed method performs better in most cases with bilinear and bicubic interpolation for  $\tilde{\mathbf{X}}$ . Obviously, bicubic interpolation is not the best upsampling approach in Fig. 6. But it shows relatively stable performance for different blur kernels and data sets. Thus, we tend to choose bicubic interpolation to obtain  $\tilde{\mathbf{X}}$ .

### C. Comparison With Other Fusion Methods

In this section, on the basis of the scenarios in Section IV-A, we compare our approaches with four state-of-the-art approaches, including the coupled nonnegative MF (CNMF) method [2], [30], the HySure method [28], the Bayesian Naive (BN) method [56], and the Bayesian Sparse (BS) method [56].

The CNMF method is first proposed in [2], which requires the relative spectral matrix and the PSF to regulate the factor matrices during the iteration process. The corresponding estimation approach for these two sensor characteristics is proposed in [30], based on which the CNMF method shows a relatively good performance for data fusion of EO-1/Hyperion and Terra/ASTER.

The HySure method models the fusion problem under the optimization framework. Subspace-based total variation (TV) regularization is introduced to impose constraints on the spectra and texture of the fused image. In this method, the spatial blur and the spectral response have also to be estimated [28], and the spatial blur is assumed to be spatially invariant, such that FFT can be used to reduce the optimization complexity. The BN method and the BS method are proposed by Wei *et al.* [6], [18]. These two methods have high order of

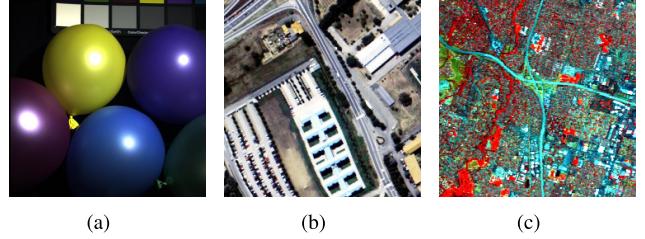


Fig. 7. (a)–(c) Ground truths of the Balloons, PaviaU, and Moffett data sets. They are all in false color.

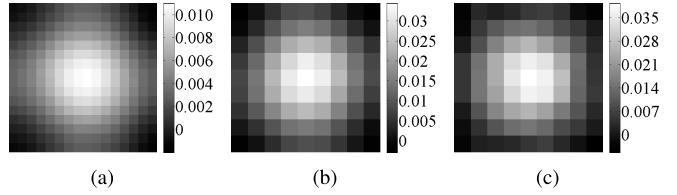


Fig. 8. Spatial blurs estimated by the approach in [28].  $\text{SNR}_m = 40 \text{ dB}$ ,  $\text{SNR}_h = 30 \text{ dB}$ . The estimation errors ( $E_b$ ) of these three blurs are 20.31%, 10.03%, and 14.32%, respectively. (a)–(c) Balloons, PaviaU, and Moffett data sets, respectively.

TABLE II  
HS+RGB IMAGE FUSION RESULTS FOR THE BALLOONS DATA SET  
( $\text{SNR}_m = 40 \text{ dB}$ ,  $\text{SNR}_h = 30 \text{ dB}$ ,  $E_f = 29.85\%$ ,  $E_b = 20.31\%$ )

Methods	RMSE	ERGAS	SAM	SSIM	Time
CNMF [2]	692.31	0.9545	6.862	0.9849	<b>34</b>
HySure [28]	511.91	0.6797	4.614	0.9922	208
BN [56]	425.32	0.5600	4.928	0.9819	<b>1</b>
BS [56]	425.80	0.5601	4.944	0.9811	2175
Proposed ( $\hat{\mathbf{F}}$ )*	<b>396.27</b>	<b>0.5241</b>	<b>3.507</b>	<b>0.9925</b>	35
Proposed ( $\mathbf{F}$ )*	<b>360.64</b>	<b>0.4782</b>	<b>3.004</b>	<b>0.9939</b>	35

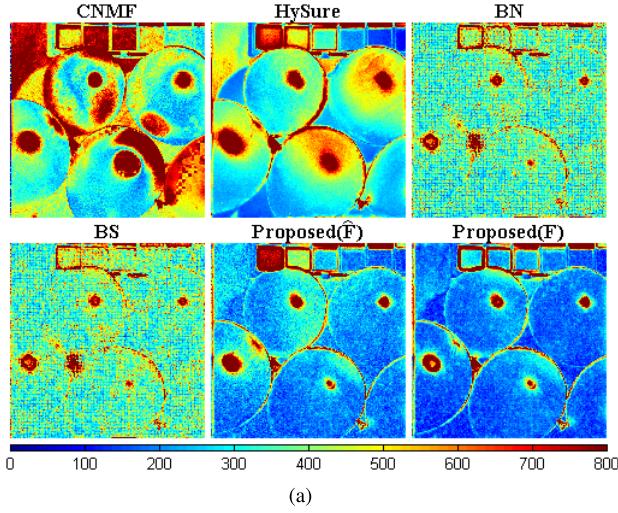
\*“Proposed( $\hat{\mathbf{F}}$ )” represents the results of our approach with the spectral responses estimated by the method in [28], while “Proposed( $\mathbf{F}$ )” represents the results of our approach with the accurate spectral responses.

complexity, especially the former one, in which Gibbs sampler is used for MMSE estimator. But later the author finds a solution to largely cut down the computational complexity for the spatially invariant imaging system [56]. Since the spatial blur and the spectral response are also required in these two methods, the author then analyzes the performance of these two methods with estimated sensor characteristics in [29]. In this paper, according to [29], we also use the approach in [28] to estimate the spatial blur and the spectral response for the BN and BS methods.

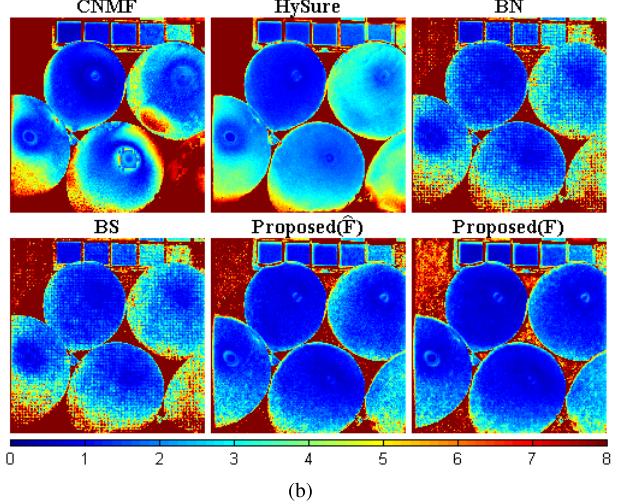
TABLE III  
HS+MS IMAGE FUSION RESULTS FOR THE PAVIAU DATA SET

Methods	SNR <sub>m</sub> = 40dB, SNR <sub>h</sub> = 30dB				SNR <sub>m</sub> = 30dB, SNR <sub>h</sub> = 20dB				SNR <sub>m</sub> = 20dB, SNR <sub>h</sub> = 20dB				Avr. Time	
	$E_f = 51.76\%, E_b = 10.03\%$				$E_f = 50.38\%, E_b = 10.11\%$				$E_f = 52.01\%, E_b = 9.57\%$					
	RMSE	ERGAS	SAM	SSIM	RMSE	ERGAS	SAM	SSIM	RMSE	ERGAS	SAM	SSIM		
CNMF [2]	171.34	2.7041	4.060	0.9679	200.20	3.0773	4.672	0.9547	274.31	4.2189	6.202	0.9042	<b>12</b>	
HySure [28]	<b>88.81</b>	1.3219	<b>2.594</b>	<b>0.9909</b>	111.76	<b>1.5499</b>	3.527	0.9850	179.57	2.3436	5.337	0.9573	42	
BN [56]	95.89	1.2642	2.761	0.9904	188.36	2.1247	6.245	0.9701	385.00	4.2949	12.445	0.8948	<b>1</b>	
BS [56]	95.70	<b>1.2591</b>	2.762	0.9905	189.22	2.1311	6.274	0.9699	385.57	4.2996	12.467	0.8947	229	
Proposed ( $\hat{\mathbf{F}}$ )*	94.83	1.4269	2.702	0.9896	<b>110.82</b>	1.5979	<b>3.329</b>	<b>0.9852</b>	<b>175.83</b>	<b>2.3375</b>	<b>4.867</b>	<b>0.9603</b>	14	
Proposed ( $\mathbf{F}$ )*	<b>64.83</b>	<b>0.9320</b>	<b>1.968</b>	<b>0.9947</b>	<b>88.08</b>	<b>1.2055</b>	<b>2.833</b>	<b>0.9900</b>	<b>164.54</b>	<b>2.1268</b>	<b>4.756</b>	<b>0.9644</b>	14	

\*“Proposed( $\hat{\mathbf{F}}$ )” represents the results of our approach with the spectral responses estimated by the method in [28], while “Proposed( $\mathbf{F}$ )” represents the results of our approach with the accurate spectral responses.



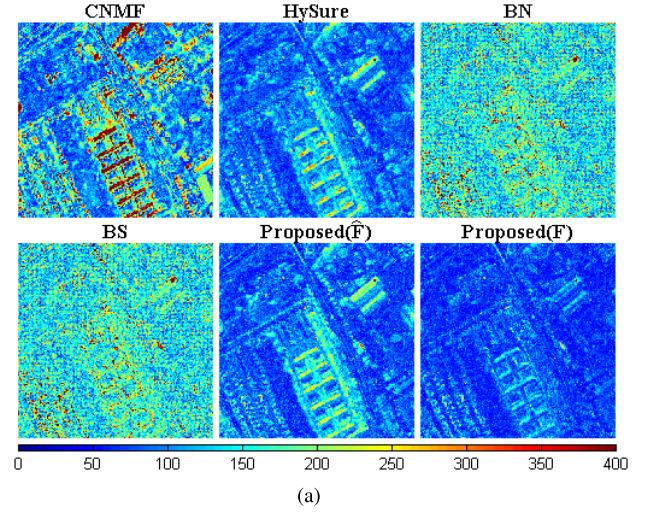
(a)



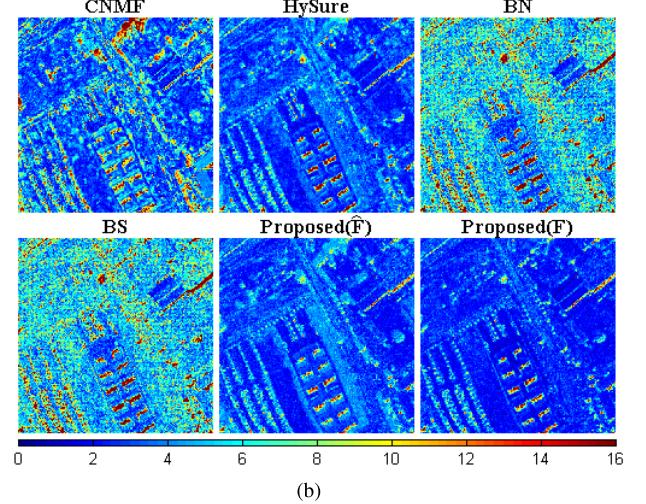
(b)

Fig. 9. Results for the Balloons data set (SNR<sub>m</sub> = 40 dB, SNR<sub>h</sub> = 30 dB). (a) RMSE images. (b) SAM images.

For a fair comparison, model parameters in the above comparison methods are tuned carefully according to their relevant papers [2], [28], [56]. Specifically, we set  $I_{in} = 200$ ,  $I_{out} = 3$  for the CNMF method with the help of Dr. Yokoya, one of the author of [2]. According to [28], we set  $\lambda_\phi = 5 \times 10^{-4}$ ,  $\lambda_m = 1$  and obtain  $\mathbf{H}$  by VCA [41] for the HySure method. As for the BN and BS methods [56], we set  $\mathbf{\Lambda}_L = \mathbf{I}_L$ ,  $\mathbf{\Lambda}_R = \mathbf{I}_L$  and obtain  $\mathbf{H}$  by truncated. Since the ranks of our fusion results ( $\mathbf{Z}$ ) are  $\min(d, r) = 10$  ( $d = 10, r = 30$ ), we set



(a)



(b)

Fig. 10. Results for the PaviaU data set (SNR<sub>m</sub> = 30 dB, SNR<sub>h</sub> = 20 dB). (a) RMSE images. (b) SAM images.

the endmembers of all above comparison methods to 10, so that the results of these four methods have the same ranks of matrices as that of our approach. Besides, to ensure the good recovery of sensor characteristics, we first define the following error assessment metrics:

$$E_f = \|\hat{\mathbf{F}} - \mathbf{F}\| / \|\mathbf{F}\| \times 100\% \quad (27)$$

$$E_b = \|\hat{\mathbf{b}} - \mathbf{b}\| / \|\mathbf{b}\| \times 100\% \quad (28)$$

TABLE IV  
HS+MS IMAGE FUSION RESULTS FOR THE MOFFETT DATA SET

Methods	SNR <sub>m</sub> = 40dB, SNR <sub>h</sub> = 30dB				SNR <sub>m</sub> = 30dB, SNR <sub>h</sub> = 20dB				SNR <sub>m</sub> = 20dB, SNR <sub>h</sub> = 20dB				Avr. Time	
	E <sub>f</sub> =98.54%, E <sub>b</sub> =14.32%				E <sub>f</sub> =99.11%, E <sub>b</sub> =12.23%				E <sub>f</sub> =99.02%, E <sub>b</sub> =9.85%					
	RMSE	ERGAS	SAM	SSIM	RMSE	ERGAS	SAM	SSIM	RMSE	ERGAS	SAM	SSIM		
CNMF [2]	107.35	2.5277	3.037	0.9224	125.38	2.6139	3.919	0.9085	172.09	<b>3.0628</b>	5.269	0.8652	<b>20</b>	
HySure [28]	88.95	2.1604	2.915	0.9472	104.29	<b>2.3903</b>	3.516	<b>0.9308</b>	156.63	<b>2.9023</b>	5.040	<b>0.8837</b>	43	
BN [56]	97.56	<b>2.0145</b>	3.011	<b>0.9479</b>	240.12	2.9124	8.417	0.8947	780.17	7.2499	23.898	0.7246	<b>1</b>	
BS [56]	97.65	<b>2.0151</b>	3.013	0.9478	240.77	2.9160	8.434	0.8945	783.20	7.2729	23.967	0.7241	286	
Proposed ( $\hat{\mathbf{F}}$ )*	<b>86.83</b>	2.1991	<b>2.891</b>	0.9464	<b>101.84</b>	<b>2.4770</b>	<b>3.506</b>	0.9280	<b>149.41</b>	3.2095	<b>4.759</b>	0.8601	25	
Proposed ( $\mathbf{F}$ )*	<b>79.41</b>	2.2167	<b>2.751</b>	<b>0.9488</b>	<b>99.73</b>	2.5005	<b>3.508</b>	<b>0.9297</b>	<b>146.32</b>	3.1556	<b>4.759</b>	<b>0.8706</b>	25	

\*“Proposed( $\hat{\mathbf{F}}$ )” represents the results of our approach with the spectral responses estimated by the method in [28], while “Proposed( $\mathbf{F}$ )” represents the results of our approach with the accurate spectral responses.

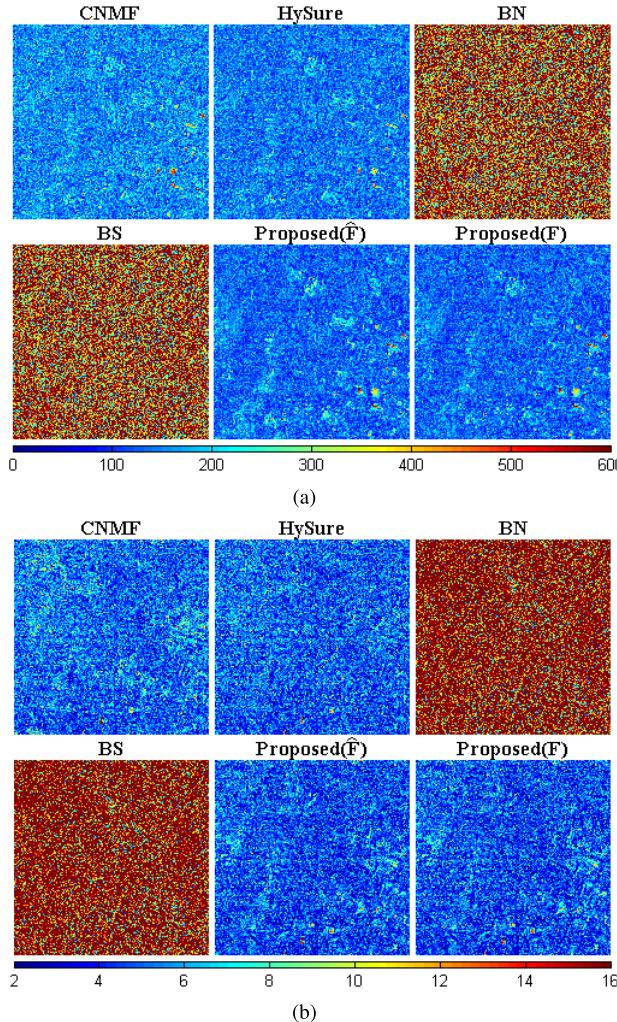


Fig. 11. Results for the Moffett data set (SNR<sub>m</sub> = 20 dB, SNR<sub>h</sub> = 20 dB). (a) RMSE images. (b) SAM images.

in which  $\hat{\mathbf{F}}$  and  $\mathbf{F}$  denote the estimated and accurate spectral response matrices, respectively, while  $\hat{\mathbf{b}}$  and  $\mathbf{b}$  denote the estimated and accurate spatial blurs, respectively. Then, we chose the values of  $\lambda_b$  and  $\lambda_R$  in the estimation approach in [28] based on these metrics. We finally set  $\lambda_b = 10$ ,  $\lambda_R = 50$  for the Balloons data set,  $\lambda_b = 100$ ,  $\lambda_R = 130$  for the PaviaU data set, and  $\lambda_b = 10$ ,  $\lambda_R = 10$  for the

Moffett data set. On the basis of the above parameter settings, we show the comparison fusion results of three data sets in Tables II–IV and Figs. 7–11, in which the bold fonts are used to show off the best two values of each quality assessment metric.

From the results of three data sets, we can find that the proposed approach with the accurate spectral response ( $\mathbf{F}$ ) surpasses others in most cases, while the proposed approach with the estimated spectral response ( $\hat{\mathbf{F}}$ ) outperforms others when the spatial blurs are not well recovered (i.e., the values of  $E_b$  are relatively large, e.g., the estimated blurs of the Balloons and the Moffett data sets) or the images are measured at low SNRs (e.g., the PaviaU data set at SNR<sub>m</sub> = 20 dB and SNR<sub>h</sub> = 20 dB). Thanks to the proposed model without intervention of the PSF, our approach is not influenced by the estimation errors of the PSF, showing more robustness to the noise compared with other fusion approaches.

We also report different fusion qualities for the comparison approaches under various SNRs. The CNMF method has poor fusion qualities in most cases despite that it costs relatively low-computational resource and is stable even at low SNR. The HySure method is the most robust one to the noise in the four comparison approaches. This method shows high SSIM values and low ERGAS values even when the SNRs turn worse. Thanks to the TV regularization, the fused images have clear texture in the PaviaU and Moffett data sets. As the parameters of the HySure method are carefully tuned according to the PaviaU data set at high SNRs in [28], the HySure method outperforms the proposed approach with  $\hat{\mathbf{F}}$  at SNR<sub>m</sub> = 40 dB and SNR<sub>h</sub> = 30 dB. Unfortunately, this method tends to be worse than the proposed when the data sets or the SNRs are changed, and it is influenced by the estimation errors of the PSF (i.e., the spatial blur). Furthermore, this method has a high computational cost, especially when the input MSI is large, e.g., the Balloons data set. As for the BN and BS methods, previous studies have proved that they show excellent performance under the condition of known sensor characteristics [32], [56]. However, in our experiment, sensor parameters are estimated instead. Although HySure, BN, and BS share the same estimated parameters, the latter two methods have poor performance at low SNR, since they are both sensitive to the estimation errors and require the manual tuning of covariance matrices,  $\Lambda_L$  and  $\Lambda_R$ , once the SNRs are changed. Moreover, we find that these two methods

are unstable for the third data set when  $\text{SNR}_m = 20$  dB and  $\text{SNR}_h = 20$  dB.

## V. CONCLUSION

In this paper, we have proposed a novel fusion model for imaging system with an unknown arbitrary PSF, which decomposes the fused image into double MF-based components and the residuals without intervention of PSF. We cast the fusion problem as an MMSE estimator under Bayesian framework. Then, a variational inference approach is utilized to obtain the posterior efficiently. The experimental results verify that our fusion approach outperforms other state-of-the-art methods, and model parameters are less dependent on the input data sets. Furthermore, as the PSF is not required in our model, estimation errors of PSF have no influence on the fusion performance. Thus, our approach shows more robustness to the noise and can be adopted for spatially variant imaging systems with no sacrifice in computational complexity.

## ACKNOWLEDGMENT

The authors would like to thank Dr. Yokoya for providing the source codes of [2] used in the comparisons and for providing help for this paper.

## REFERENCES

- [1] C. Ratto, K. D. Morton, L. M. Collins, and P. A. Torrione, "Bayesian context-dependent learning for anomaly classification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 4, pp. 1969–1981, Apr. 2014.
- [2] 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.
- [3] N. Akhtar, F. Shafait, and A. Mian, "Bayesian sparse representation for hyperspectral image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3631–3640.
- [4] 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.
- [5] 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.
- [6] Q. Wei, J. Bioucas-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.
- [7] C. Chen, Y. Li, W. Liu, and J. Huang, "Sirf: Simultaneous satellite image registration and fusion in a unified framework," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4213–4224, Nov. 2015.
- [8] I. Amro, J. Mateos, M. Vega, R. Molina, and A. K. Katsaggelos, "A survey of classical methods and new trends in pansharpening of multispectral images," *EURASIP J. Adv. Signal Process.*, vol. 2011, no. 1, p. 1, 2011. [Online]. Available: <https://link.springer.com/article/10.1186/1687-6180-2011-79>
- [9] I. Jolliffe, *Principal Component Analysis*. Hoboken, NJ, USA: Wiley, 2002.
- [10] G. Licciardi, M. Khan, and J. Chanussot, "Fusion of hyperspectral and panchromatic images: A hybrid use of indusion and nonlinear PCA," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2012, pp. 2133–2136.
- [11] V. P. Shah, N. H. Younan, and R. L. King, "An efficient pan-sharpening method via a combined adaptive PCA approach and contourlets," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1323–1335, May 2008.
- [12] M. Gonzalez-Audicana, J. L. Saleta, R. G. Catalán, and R. García, "Fusion of multispectral and panchromatic images using improved IHS and PCA mergers based on wavelet decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 6, pp. 1291–1299, Jun. 2004.
- [13] C. Thomas, T. Ranchin, L. Wald, and J. Chanussot, "Synthesis of multispectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1301–1312, May 2008.
- [14] Y. Zhao, C. Yi, J. Yang, and J. C. Chan, "Coupled hyperspectral super-resolution and unmixing," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Sep. 2014, pp. 2641–2644.
- [15] C. Chen, Y. Li, W. Liu, and J. Huang, "Image fusion with local spectral consistency and dynamic gradient sparsity," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 2760–2765.
- [16] M. A. Veganzones, M. Simões, G. Licciardi, J. Bioucas, and J. Chanussot, "Hyperspectral super-resolution of locally low rank images from complementary multisource data," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2014, pp. 703–707.
- [17] O. Eches, N. Dobigeon, J. Tourneret, and H. Snoussi, "Variational methods for spectral unmixing of hyperspectral images," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Sep. 2011, pp. 957–960.
- [18] 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.
- [19] R. Kawakami, J. Wright, Y. Tai, Y. Matsushita, M. Ben-Ezra, and K. Ikeuchi, "High-resolution hyperspectral imaging via matrix factorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Sep. 2011, pp. 2329–2336.
- [20] M. A. Bendoumi, M. He, and S. Mei, "Hyperspectral image resolution enhancement using high-resolution multispectral image based on spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6574–6583, Oct. 2014.
- [21] S. Li, H. Yin, and L. Fang, "Remote sensing image fusion via sparse representations over learned dictionaries," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4779–4789, Sep. 2013.
- [22] E. Wycoff, T. Chan, K. Jia, W. Ma, and Y. Ma, "A non-negative sparse promoting algorithm for high resolution hyperspectral imaging," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Apr. 2013, pp. 1409–1413.
- [23] N. Akhtar, F. Shafait, and A. Mian, "Sparse spatio-spectral representation for hyperspectral image super-resolution," in *Computer Vision—ECCV 2014*. Switzerland: Springer, 2014, pp. 63–78.
- [24] X. Liu, W. Xia, B. Wang, and L. Zhang, "An approach based on constrained nonnegative matrix factorization to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 757–772, Feb. 2011.
- [25] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.
- [26] K. Rong, L. Jiao, S. Wang, and F. Liu, "Pansharpening based on low-rank and sparse decomposition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 12, pp. 4793–4805, Dec. 2014.
- [27] H. B. Mitchell, *Image Fusion: Theories, Techniques and Applications*. Berlin, Germany: Springer-Verlag, 2010.
- [28] M. Simões, 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.
- [29] Q. Wei, J. Bioucas-Dias, N. Dobigeon, J. Y. Tourneret, and S. Godsill, "Blind model-based fusion of multi-band and panchromatic images," in *Proc. IEEE Int. Conf. Multisensor Fusion Integr. Intell. Syst.*, 2016, pp. 1–5.
- [30] N. Yokoya, N. Mayumi, and A. Iwasaki, "Cross-calibration for data fusion of EO-1/hyperion and Terra/ASTER," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 419–426, Apr. 2013.
- [31] J. Bardsley, S. Jefferies, J. Nagy, and R. Plemmens, "A computational method for the restoration of images with an unknown, spatially-varying blur," *Opt. Exp.*, vol. 14, no. 5, pp. 1767–1782, 2006.
- [32] L. Loncan *et al.*, "Hyperspectral pansharpening: A review," *IEEE Trans. Geosci. Remote Sens.*, vol. 3, no. 3, pp. 27–46, Sep. 2015.
- [33] B. Lin, X. Tao, L. Dong, and J. Lu, "Variational em approach for high resolution hyper-spectral imaging based on probabilistic matrix factorization," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 1774–1778.
- [34] B. Lin, X. Tao, S. Li, L. Dong, and J. Lu, "Variational Bayesian image fusion based on combined sparse representations," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Sep. 2016, pp. 1432–1436.

- [35] M. Combescure, "Block-circulant matrices with circulant blocks, Weil sums, and mutually unbiased bases. II. The prime power case," *J. Math. Phys.*, vol. 50, no. 3, p. 032104, 2009.
- [36] J. G. Nagy and D. P. O'leary, "Fast iterative image restoration with a spatially varying PSF," *Proc. SPIE*, vol. 3162, p. 388, Oct. 1997. [Online]. Available: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=930371>
- [37] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. 25, no. 9, pp. 772–781, Sep. 1978.
- [38] A. Mnih and R. Salakhutdinov, "Probabilistic matrix factorization," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2007, pp. 1257–1264.
- [39] W. He, H. Zhang, L. Zhang, and H. Shen, "Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 176–188, Jan. 2016.
- [40] G. W. Stewart, "On the early history of the singular value decomposition," *SIAM Rev.*, vol. 35, no. 4, pp. 551–566, 1993.
- [41] J. M. P. Nascimento and J. M. Bioucas-Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.
- [42] J. H. Wilkinson, F. L. Bauer, and C. Reinsch, *Linear Algebra*, vol. 2. New York, NY, USA: Springer-Verlag, 2013,
- [43] X. Ding, L. He, and L. Carin, "Bayesian robust principal component analysis," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3419–3430, May 2011.
- [44] R. Salakhutdinov and A. Mnih, "Bayesian probabilistic matrix factorization using Markov Chain Monte Carlo," in *Proc. IEEE Int. Conf. Mach. Learn. (ICML)*, Sep. 2008, pp. 880–887.
- [45] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 131–146, Nov. 2008.
- [46] S. Li, X. Tao, and J. Lu, "A variational Bayesian em approach to structured sparse signal reconstruction," in *Proc. IEEE 80th Veh. Technol. Conf. (VTC Fall)*, Oct. 2014, pp. 1–5.
- [47] S. D. Babacan, M. Luessi, R. Molina, and A. K. Katsaggelos, "Sparse Bayesian methods for low-rank matrix estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 3964–3977, Aug. 2012.
- [48] G. H. Golub and C. F. Van Loan, *Matrix Computations*, vol. 3. Baltimore, MD, USA: JHU Press, 2012,
- [49] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images," *Photogramm. Eng. Remote Sens.*, vol. 63, no. 6, pp. 691–699, 1997.
- [50] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2241–2253, Sep. 2010.
- [51] B. Huang, H. Song, H. Cui, J. Peng, and Z. Xu, "Spatial and spectral image fusion using sparse matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 3, pp. 1693–1704, Mar. 2014.
- [52] L. Wald, "Quality of high resolution synthesised images: Is there a simple criterion?" in *Proc. 3rd Conf. Fusion Earth Data, Merging Point Meas., Raster Maps Remote. Sensed Images*, 2000, pp. 99–103.
- [53] 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, Apr. 2004.
- [54] G. Casella and E. I. George, "Explaining the Gibbs sampler," *Amer. Statist.*, vol. 46, no. 3, pp. 167–174, 1992.
- [55] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sensing*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.
- [56] Q. Wei, N. Dobigeon, and J. 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.



**Baihong Lin** received the B.S. degree from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 2013. He is currently pursuing the Ph.D. degree with the Department of Electronic Engineering, Tsinghua University, Beijing.

His research interests include image processing, medical imaging, computer vision, and machine learning.



**Xiaoming Tao** (M'09) received the B.E. degree from the School of Telecommunications Engineering, Xidian University in 2003, and Ph.D. from the Department of Electronic Engineering, Tsinghua University, Beijing, China, in 2008.

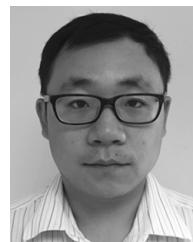
She is currently an Associate Professor with the Department of Electronic Engineering, Tsinghua University. Her research interests include wireless communication and networking, and multimedia signal processing.



**Mai Xu** (M'10–SM'17) received the B.S. degree from Beihang University, Beijing, China, in 2003, the M.S. degree from Tsinghua University, Beijing, China, in 2006, and the Ph.D. degree from Imperial College London, London, U.K., in 2010.

From 2010 to 2012, he was a Research Fellow with the Electrical Engineering Department, Tsinghua University. From 2014 to 2015, he was a Visiting Researcher with MSRA. Since 2013, he has been with Beihang University as an Associate Professor. He has authored over 70 technical papers in international journals and conference proceedings. His research interests include visual communication and image processing.

Dr. Xu was a recipient of best paper awards of two IEEE conferences.



**Linhao Dong** (S'13–M'15) received the B.E. degree from Dalian Nationalities University, Dalian, China, the B.Eng. degree from Glyndwr University, Wrexham, U.K., in 2009, and the Ph.D. degree from the University of Liverpool, Liverpool, U.K., in 2014.

He was a Post-Doctoral Researcher with the School of Aerospace Engineering, Tsinghua University, Beijing, China, from 2014 to 2016. He has been a Post-Doctoral Researcher with the College of Information and Communication Engineering, Sungkyunkwan University, Suwon, South Korea, since 2016. His research interests include green communications, wireless resource allocation, cooperative networks, millimeter-wave communications, energy harvesting communications, and wireless multimedia communications.



**Jianhua Lu** (M'98–SM'07–F'15) received the B.S.E.E. and M.S.E.E. degrees from Tsinghua University, Beijing, China, in 1986 and 1989, respectively, and the Ph.D. degree in electrical and electronic engineering from the Hong Kong University of Science and Technology, Hong Kong.

Since 1989, he has been with the Department of Electronic Engineering, Tsinghua University, where he serves as a Professor. He has authored more than 180 technical papers in international journals and conference proceedings. His research interests include broadband wireless communication, multimedia signal processing, and wireless networking.

Dr. Lu is also a Fellow of the IEEE Communication Society and the IEEE Signal Processing Society. He has served in numerous IEEE conferences as a member of Technical Program Committees and served as the Lead Chair of the General Symposium of IEEE ICC 2008, as well as a Program Committee Co-Chair of the Ninth IEEE International Conference on Cognitive Informatics in 2010. He has been an active member of professional societies. He was a recipient of best paper awards at the IEEE International Conference on Communications, Circuits and Systems 2002, ChinaCom 2006, and IEEE Embedded-Com 2012, and the National Distinguished Young Scholar Fund by the NSF Committee of China in 2005. He is now a Chief Scientist of the National Basic Research Program (973), China.