VARIATIONAL BAYESIAN IMAGE FUSION BASED ON COMBINED SPARSE REPRESENTATIONS

Baihong Lin, Xiaoming Tao, Shaoyang Li, Linhao Dong, and Jianhua Lu

Tsinghua National Laboratory for Information Science and Technology (TNList), Beijing, China

ABSTRACT

Hyper-spectral image fusion has been a hot topic in medical imaging and remote sensing. This paper proposes a Bayesian fusion model which combines the panchromatic (PAN) image and the low spatial resolution hyper-spectral (HS) image under the same framework. Sparsity constraint is introduced as double "spike-and-slab" priors, and anisotropic Gaussian noise is adopted for accuracy. To achieve reduction in computational complexity, we turn the anisotropic Gaussian distribution into isotropic one with modified linear transformation and propose a variational Bayesian expectation maximization (EM) algorithm to calculate the result. Experiment results show that our solution can achieve comparable performance in pan-sharpening to other state-of-art algorithms while largely reducing the computational complexity.

Index Terms— Hyper-spectral image fusion, sparse prior, computational complexity, anisotropic Gaussian distribution, variational Bayesian algorithm

1. INTRODUCTION

Hyper-spectral image fusion, also called pan-sharpening [1, 2], aims at recovering a high spatial resolution hyper-spectral image from a high spatial resolution panchromatic (PAN) image and a low spatial resolution hyper-spectral (HS) image. This technique has been a hot topic in recent years and it compensates for the technical defect of hyper-spectral sensors [3, 4], playing an important role in medical imaging, geosciences and remote sensing, especially in some occasions such as detection and material identification [1].

Recently, it is found that the high spatial resolution hyper-spectral image could be pan-sharpened by matrix factorization [3, 5]. Inspired by such finding, various methods were proposed to enhance the performance of pan-sharpening [1, 3, 6, 7, 8, 9, 10], among which bayesian sparse representation of dictionary outstrips others in terms of recovery quality with great success [4]. This Bayesian model divides the fusion process into two stages where Markov chain Monte Carlo (MCMC) algorithm is used to figure out the pan-sharpened image. However, such method is computationally complex because of long term sampling, and its convergence is hardly guaranteed in finite steps [11]. What's worse, solutions got by the method proposed in [4] are still not global optima because of the separate processing on the PAN image and the low spatial resolution hyper-spectral image.

In this paper, we first reformulate the Bayesian model of the hyper-spectral image fusion with new mathematical expressions. Different from the work in [4], our model combines two individual stages, the dictionary learning stage and the sparse coding stage, which are formulated as probability distributions under the same framework of Bayesian model. Based on mathematical analysis, double "spike-and-splab" priors and anisotropic Gaussian noise distribution are introduced to accurately describe the relationship between these two kinds of images. This model is more likely to bring in a global optimal solution, as the dictionary learning and the sparse coding are processing on the PAN image and low spatial resolution image simultaneously. Second, we adopt variational Bayesian approximation method to reduce the computational complexity [11]. Although this method introduces loss in performance, it effectively avoids long sampling time and the reconstruction quality can still be guaranteed by the unique features of our model. Experiments show that the result of our scheme is comparable in terms of reconstruction performance to other state-of-art algorithms, while largely cutting down the computational cost. For images in CAVE database, our algorithm can even run faster by 13.58 times with promising recovery quality, compared to [4].

The paper is organized as follows. Section 2 formulates our Bayesian model. Section 3 shows the detail of the variational Bayesian algorithm. Finally, the simulation results are given in Section 4 and conclusion is drawn in Section 5.

2. BAYESIAN MODEL FORMULATION

Let the PAN image and the low spatial resolution hyper-spectral image be denoted by $\mathbf{Y} \in \mathbb{R}^{l \times MN}$ and $\mathbf{X} \in \mathbb{R}^{L \times mn}$ respectively, wherein l and L denote their respective channel numbers. MN and mn are their corresponding image sizes. Then the pan-sharpened image denoted by $\mathbf{Z} \in \mathbb{R}^{L \times MN}$

This work was supported by the National Basic Research Project of China (973)(2013CB329006), National Natural Science Foundation of China (NSFC, 61471220, 61021001), and Tsinghua University Initiative Scientific Research Program.

satisfies the following equations, as:

$$\mathbf{Y} = \mathbf{FZ}, \text{ and } \mathbf{X} = \mathbf{ZG}$$

wherein $\mathbf{F} \in \mathbb{R}^{l\times L}$ denotes the spectral transform matrix and $\mathbf{G} \in \mathbb{R}^{MN\times mn}$ denotes the spatial transform matrix. According to the linear spectral mixture model [3], the pan-sharpened image could be expressed as following:

$$\mathbf{Z} = \mathbf{AT} + \mathbf{N}$$

wherein $\mathbf{A}$ is defined as a spectral dictionary and $\mathbf{T}$ is the related sparse representation, with $\mathbf{N}$ denoting the residual. We assume that the residual $\mathbf{N}$ is an isotropic Gaussian noise as $n_{(i,j)} \sim \mathcal{N}(0, \alpha_{\mathbf{N}}^{-1})$. Then the distributions of $\mathbf{Y}$ are as following, wherein $\mathbf{Y}_{.i}$ denotes the i^{th} column of the matrix $\mathbf{Y}$.

$$\mathbf{Y}|\mathbf{A},\mathbf{T} \sim \prod_{i=1}^{MN} \mathcal{N}(\mathbf{Y}_{.i}|\mathbf{FAT}_{.i}, \alpha_{\mathbf{N}}^{-1}\mathbf{FF}^T)$$

which is an anisotropic Gaussian distribution. To simplify the computation, we transform $\mathbf{Y}$ into $\widetilde{\mathbf{Y}} = \mathbf{QY}$,wherein $\mathbf{Q} \in \mathbb{R}^{l\times l}$, $\mathbf{Q}^T\mathbf{Q} = (\mathbf{FF}^T)^{-1}$, and let $\widetilde{\mathbf{F}} = \mathbf{QF}$. Apparently, $\mathbf{Q}$ could be easily yielded from SVD. For simplicity, we drop $\widetilde{\mathbf{F}}$ and $\widetilde{\mathbf{Y}}$ to $\mathbf{F}$ and $\mathbf{Y}$, respectively, in the following part of this paper. Then we have

$$\mathbf{Y}|\mathbf{A},\mathbf{T} \sim \prod_{i=1}^{MN} \mathcal{N}(\mathbf{Y}_{.i}|\mathbf{FAT}_{.i}, \alpha_{\mathbf{N}}^{-1}\mathbf{I}_l)$$

As for the distribution of $\mathbf{X}$, we only discuss the case that the matrix $\mathbf{G}$ is unknown, given $\mathbf{S} = \mathbf{TG}$. To further simplify the problem, we assume that $\mathbf{S}$ is independent from $\mathbf{T}$ and the residual of $\mathbf{NG}$ is also an isotropic Gaussian noise. Thus the distribution is:

$$\mathbf{X}|\mathbf{A},\mathbf{S} \sim \prod_{j=1}^{mn} \mathcal{N}(\mathbf{X}_{.j}|\mathbf{AS}_{.j}, \alpha_{\mathbf{X}}^{-1}\mathbf{I}_L)$$

To depict the sparsity, we adopt the "spike-and-slab" priors for $\mathbf{S}$ and $\mathbf{T}$ [12]. Dictionary prior and hyper-parameter conjugate prior are succeeded from the work [4]. Whereas in this paper, we put these two sparse matrices under the same Bayesian framework, which guarantees strict mathematical logic. The priors are shown as follows, wherein $\circ$ denotes Hadamard product [13].

$$\mathbf{A} \sim \prod_{k=1}^{K} \mathcal{N}(\mu_{\mathbf{A}_{k0}}, \alpha_{\mathbf{A}_{k0}}^{-1}\mathbf{I}_L)$$

$$\mathbf{S} = \mathbf{W}\circ\mathbf{\Theta},\ \mathbf{W}_{i,j} \sim Bernoulli(\beta_i),\ \mathbf{\Theta}_{ij} \sim \mathcal{N}(0, \alpha_{\mathbf{\Theta}}^{-1})$$
$$\mathbf{T} = \mathbf{\Pi}\circ\mathbf{B},\ \mathbf{\Pi}_{i,j} \sim Bernoulli(\gamma_i),\ \mathbf{B}_{ij} \sim \mathcal{N}(0, \alpha_{\mathbf{B}}^{-1})$$
$$\alpha_{\mathbf{X}}(\alpha_{\mathbf{N}}, \alpha_{\mathbf{\Theta}}, \alpha_{\mathbf{B}}) \sim Gamma(a_{\mathbf{X}(\mathbf{N},\mathbf{\Theta},\mathbf{B})}, b_{\mathbf{X}(\mathbf{N},\mathbf{\Theta},\mathbf{B})})$$
$$\beta_i(\gamma_i) \sim Beta(c_{\beta(\gamma)}/K, d_{\beta(\gamma)}(K-1)/K)$$

Based on the assumption and the prior distribution above, we could get the joint distribution of the hierarchical Bayesian model.

$$\begin{aligned} &p(\mathbf{X},\mathbf{Y},\mathbf{A},\mathbf{S},\mathbf{T},\alpha_{\mathbf{X}},\alpha_{\mathbf{N}}) \\ &\quad = p(\mathbf{Y}|\mathbf{A},\mathbf{T},\alpha_{\mathbf{N}})p(\alpha_{\mathbf{N}})p(\mathbf{X}|\mathbf{A},\mathbf{S},\alpha_{\mathbf{X}})p(\mathbf{A})p(\alpha_{\mathbf{X}}) \\ &\quad p(\mathbf{W}|\beta)p(\beta)p(\mathbf{\Theta}|\alpha_{\mathbf{\Theta}})p(\mathbf{\Pi}|\gamma)p(\gamma)p(\mathbf{B}|\alpha_{\mathbf{B}})p(\alpha_{\mathbf{B}}) \end{aligned}$$

3. VARIATIONAL ALGORITHM

Apparently, no closed solution exists for our Bayesian model introduced in Section 2. Thus approximation needs to be made. Although MCMC could be used to figure out the posteriori distribution of the model, such algorithm requires great computational complexity because of the long term sampling process that guaranteeing the convergence to the solution [11, 14]. To reduce the time consumption, we hereby propose a variational Bayesian EM algorithm.

In our algorithm, the posteriori distribution is factorized approximately, which is:

$$\begin{aligned} &p(\mathbf{A},\mathbf{S},\mathbf{T},\alpha_{\mathbf{X}},\alpha_{\mathbf{N}}|\mathbf{X},\mathbf{Y}) \approx q(\mathbf{\Theta})q(\mathbf{W})q(\mathbf{B})q(\mathbf{\Pi}) \\ &\quad \prod_{k=1}^{K} q(\mathbf{A}_{.k})q(\alpha_{\mathbf{X}})q(\alpha_{\mathbf{N}})q(\alpha_{\mathbf{\Theta}})q(\alpha_{\mathbf{B}}) \prod_{i=1}^{K} q(\beta_i) \prod_{i=1}^{K} q(\gamma_i) \end{aligned}$$

wherein $q(.)$ denotes posterior approximation of the corresponding variable. Under the variational framework, we iterate to minimize the Kullback-Leibler distance between the posteriori distribution and the approximation distribution [11]. Let $\mathcal{P} = \{\mathbf{\Theta},\mathbf{W},\mathbf{B},\mathbf{\Pi},\mathbf{A}_{.k},\alpha_{\mathbf{X}},\alpha_{\mathbf{N}},\alpha_{\mathbf{\Theta}},\alpha_{\mathbf{B}},\beta_i,\gamma_i\}$. For each $\mathbf{p}_k \,\epsilon\, \mathcal{P}$, the posterior approximation could be updated by the following equation, wherein $< . >_{\mathbf{A}}$ denotes the corresponding expectation under the distribution of $\mathbf{A}$ and $\mathcal{P} \setminus \mathbf{p}_k$ denotes all variables in $\mathcal{P}$ except $\mathbf{p}_k$.

$$\ln q(\mathbf{p}_k) = < \ln p(\mathbf{A},\mathbf{S},\mathbf{T},\alpha_{\mathbf{X}},\alpha_{\mathbf{N}}|\mathbf{X},\mathbf{Y}) >_{\mathcal{P}\setminus\mathbf{p}_k} + \text{const}$$

3.1. Posterior Approximation of W and Π

According to the factorized joint distribution, the approximated posterior of $\mathbf{W}$ could be derived as following:

$$\begin{aligned} q(\mathbf{W}_{i,j}=1) \propto \exp\{-\frac{1}{2}[\overline{\alpha}_{\mathbf{X}}\overline{\mathbf{A}^T\mathbf{A}}_{(i,i)}(\mu^2_{\mathbf{\Theta}_{i,j}} + \mathbf{\Sigma}_{\mathbf{\Theta}_{.j}(i,i)}) \\ -2\overline{\alpha}_{\mathbf{X}}\mu_{\mathbf{\Theta}_{i,j}}(\overline{\mathbf{A}}^T_{.i}\mathbf{X}_{.j} - \overline{\mathbf{A}^T\mathbf{A}}_{(i,(-i))}\overline{\mathbf{S}}_{(-i),j})] + \overline{\ln\beta_{i,j}}\} \end{aligned} \tag{1}$$

$$\overline{\mathbf{W}}_{i,j} = \frac{q(\mathbf{W}_{i,j}=1)}{q(\mathbf{W}_{i,j}=1) + \exp(\overline{\ln(1-\beta_{i,j})})} \tag{2}$$

where $\overline{(.)}$ denotes the statistical mean and $\mathbf{A}_{(-i)j}$ denotes the j^{th} column of $\mathbf{A}$ except the i^{th} row. Similar to $\mathbf{W}$, $\mathbf{\Pi}$ has the same type of distribution and can be obtained by replacing $\mathbf{X}$, $\mathbf{S}$, $\mathbf{A}$ and $\alpha_{\mathbf{X}}$ with $\mathbf{Y}$, $\mathbf{T}$, $\mathbf{FA}$ and $\alpha_{\mathbf{N}}$ in eq.(1) and eq.(2), respectively.

3.2. Posterior Approximation of Θ and B

The distribution of $\boldsymbol{\Theta}$ can be derived as a Gaussian distribution, and its parameters are determined by the following equation.

$$q(\boldsymbol{\Theta}) \sim \prod_{j=1}^{mn} \mathcal{N}(\mu_{\Theta_{.j}}, \boldsymbol{\Lambda}_{\Theta_{.j}}^{-1}) \tag{3}$$

$$\mu_{\boldsymbol{\Theta}_{.j}} = \overline{\alpha}_{\mathbf{X}} \boldsymbol{\Lambda}_{\boldsymbol{\Theta}_{.j}}^{-1} \mathrm{diag}(\overline{\mathbf{W}}_{.j}) \overline{\mathbf{A}}^T \mathbf{X}_{.j}$$

$$\boldsymbol{\Lambda}_{\boldsymbol{\Theta}_{.j}} = \overline{\alpha}_{\mathbf{X}} \overline{\mathbf{A}^T\mathbf{A}} \circ \overline{\mathbf{W}_{.j}\mathbf{W}_{.j}^T} + \overline{\alpha}_{\boldsymbol{\Theta}} \mathbf{I}_K$$

$$\overline{\mathbf{W}_{.j}\mathbf{W}_{.j}^T} = \overline{\mathbf{W}}_{.j} \overline{\mathbf{W}}_{.j}^T + \mathrm{diag}[\overline{\mathbf{W}}_{.j} \circ (1 - \overline{\mathbf{W}}_{.j})]$$

Additionally, $\overline{\boldsymbol{\Theta}\boldsymbol{\Theta}^T}$ and $\overline{\mathbf{S}\mathbf{S}^T}$ ($\overline{\mathbf{B}\mathbf{B}^T}$ and $\overline{\mathbf{T}\mathbf{T}^T}$) have to be updated for other posterior approximations. As the posterior approximation is similar to that of $\boldsymbol{\Theta}$, the distribution of $\mathbf{B}$, $\overline{\mathbf{B}\mathbf{B}^T}$ and $\overline{\mathbf{T}\mathbf{T}^T}$ can be obtained by substituting $\mathbf{X}$, $\mathbf{S}$, $\mathbf{A}$, $\mathbf{W}$, $\alpha_{\mathbf{X}}$ and $\alpha_{\boldsymbol{\Theta}}$ into $\mathbf{Y}$, $\mathbf{T}$, $\mathbf{FA}$, $\boldsymbol{\Pi}$, $\alpha_{\mathbf{N}}$ and $\alpha_{\mathbf{B}}$, respectively.

3.3. Posterior Approximation of $\mathbf{A}_{.k}$

The posterior distribution of $\mathbf{A}$ is factorized into K factors, $\mathbf{A}_{.k}, k = 1, 2, ..., K$, approximately. Each factor follows a Gaussian distribution and their parameters could be obtained as follows:

$$q(\mathbf{A}_{.k}) \sim \mathcal{N}(\mu_{\mathbf{A}_{.k}}, \boldsymbol{\Lambda}_{\mathbf{A}_{.k}}^{-1}) \tag{4}$$

$$\boldsymbol{\Lambda}_{\mathbf{A}_{.k}} = (\overline{\alpha}_{\mathbf{X}} \overline{\mathbf{S}\mathbf{S}^T}_{(k,k)} + \alpha_{\mathbf{A}_{k0}}) \mathbf{I}_L + \overline{\alpha}_{\mathbf{N}} \overline{\mathbf{T}\mathbf{T}^T}_{(k,k)} \mathbf{F}^T \mathbf{F}$$

$$\begin{aligned} \mu_{\mathbf{A}_{.k}} =& \boldsymbol{\Lambda}_{\mathbf{A}_{.k}}^{-1} \{ \overline{\alpha}_{\mathbf{X}} (\mathbf{X}\overline{\mathbf{S}}_{(k.)}^T - \overline{\mathbf{A}}_{(-k)} \overline{\mathbf{S}\mathbf{S}^T}_{(-k),k}) \\ &+ \overline{\alpha}_{\mathbf{N}} \mathbf{F}^{\mathrm{T}} (\mathbf{Y}\overline{\mathbf{T}}_{(k.)}^T - \mathbf{F}\overline{\mathbf{A}}_{(-k)} \overline{\mathbf{T}\mathbf{T}^T}_{(-k),k}) \} \end{aligned}$$

$\overline{\mathbf{A}^T\mathbf{A}}$ and $\overline{\mathbf{A}^T\mathbf{F}^T\mathbf{F}\mathbf{A}}$ have to be updated for posterior approximation of $\boldsymbol{\Theta}$ and $\mathbf{W}$ ($\mathbf{B}$ and $\boldsymbol{\Pi}$). They could be easily got based on the distribution of $\mathbf{A}_{.k}$. Due to the space limit, this deriving procedure is not shown here.

3.4. Posterior Approximation of Hyper-parameters

As conjugate priors are adopted, all the posterior of hyper-parameters will hold the same type of distribution form as their corresponding prior. The approximations are as follows, wherein $\mathrm{tr}(.)$ denotes the trace of the matrix.

$$\begin{aligned} q(\alpha_{\mathbf{X}}) \sim Gamma(a_{\mathbf{X}} + \frac{Lmn}{2}, b_{\mathbf{X}} + \frac{1}{2}[\mathrm{tr}(\mathbf{X}\mathbf{X}^T) \\ - 2\mathrm{tr}(\overline{\mathbf{A}}\overline{\mathbf{S}}\mathbf{X}^T) + \mathrm{tr}(\overline{\mathbf{A}^T\mathbf{A}}\overline{\mathbf{S}\mathbf{S}^T})]) \end{aligned} \tag{5}$$

$$\begin{aligned} q(\alpha_{\mathbf{N}}) \sim Gamma(a_{\mathbf{N}} + \frac{lMN}{2}, b_{\mathbf{N}} + \frac{1}{2}[\mathrm{tr}(\mathbf{Y}\mathbf{Y}^T) \\ - 2\mathrm{tr}(\mathbf{F}\overline{\mathbf{A}}\overline{\mathbf{T}}\mathbf{Y}^T) + \mathrm{tr}(\overline{\mathbf{A}^T\mathbf{F}^T\mathbf{F}\mathbf{A}}\overline{\mathbf{T}\mathbf{T}^T})]) \end{aligned} \tag{6}$$

$$q(\alpha_{\boldsymbol{\Theta}}) \sim Gamma(a_{\boldsymbol{\Theta}} + \frac{Kmn}{2}, b_{\boldsymbol{\Theta}} + \frac{1}{2}\mathrm{tr}(\overline{\boldsymbol{\Theta}\boldsymbol{\Theta}^T})) \tag{7}$$

Algorithm 1 Variational Bayesian Algorithm

Input: $\mathbf{X}$,$\mathbf{Y}$,K,T_{stop}

Initialize: the hyper-parameters $a_{\mathbf{X}}$, $b_{\mathbf{X}}$, $a_{\mathbf{N}}$, $b_{\mathbf{N}}$, $a_{\boldsymbol{\Theta}}$, $b_{\boldsymbol{\Theta}}$, $a_{\mathbf{B}}$, $b_{\mathbf{B}}$, c_{β}, d_{β}, c_{γ}, d_{γ} and the parameters $\overline{\mathbf{A}}$, $\overline{\mathbf{A}^T\mathbf{F}^T\mathbf{F}\mathbf{A}}$, $\overline{\mathbf{A}^T\mathbf{A}}$, $\overline{\boldsymbol{\Pi}}$, $\overline{\mathbf{W}}$, $t = 0$

Output: $\mathbf{Z} = \overline{\mathbf{A}\mathbf{T}}$.

While $t \leq T_{stop}$ **do**

1. Update $\overline{\mathbf{W}}$ (and $\overline{\boldsymbol{\Pi}}$) with the Eqs. (1)~(2) if it is not the first iteration.
2. Update $\overline{\boldsymbol{\Theta}}$, $\overline{\boldsymbol{\Theta}\boldsymbol{\Theta}^T}$, $\overline{\mathbf{S}\mathbf{S}^T}$ (and $\overline{\mathbf{B}}$, $\overline{\mathbf{B}\mathbf{B}^T}$, $\overline{\mathbf{T}\mathbf{T}^T}$) with the Eq. (3).
3. Update $\overline{\mathbf{A}}$, $\overline{\mathbf{A}^T\mathbf{A}}$, $\overline{\mathbf{A}^T\mathbf{F}^T\mathbf{F}\mathbf{A}}$ with the Eq. (4)
4. Update $q(\alpha_{\mathbf{X}})$, $q(\alpha_{\mathbf{N}})$, $q(\alpha_{\boldsymbol{\Theta}})$, $q(\alpha_{\mathbf{B}})$, $q(\alpha_{\beta_i})$, $q(\alpha_{\gamma_i})$ and $\overline{\alpha}_{\mathbf{X}}$, $\overline{\alpha}_{\mathbf{N}}$, $\overline{\alpha}_{\boldsymbol{\Theta}}$, $\overline{\alpha}_{\mathbf{B}}$, $\overline{\beta}_i$, $\overline{\gamma}_i$ with Eqs. (5) $\sim$ (10)
5. $t \leftarrow t + 1$

End while

$$q(\alpha_{\mathbf{B}}) \sim Gamma(a_{\mathbf{B}} + \frac{KMN}{2}, b_{\mathbf{B}} + \frac{1}{2}\mathrm{tr}(\overline{\mathbf{B}\mathbf{B}^T})) \tag{8}$$

$$q(\beta_i) \sim Beta(c_{\beta} + \sum_{j=1}^{mn} \overline{\mathbf{W}}_{(i,j)}, d_{\beta} + \sum_{j=1}^{mn} (1 - \overline{\mathbf{W}}_{(i,j)})) \tag{9}$$

$$q(\gamma_i) \sim Beta(c_{\gamma} + \sum_{j=1}^{MN} \overline{\boldsymbol{\Pi}}_{(i,j)}, d_{\gamma} + \sum_{j=1}^{MN} (1 - \overline{\boldsymbol{\Pi}}_{(i,j)})) \tag{10}$$

Our algorithm is summarized in Algorithm 1. As $\overline{\mathbf{W}}$ (and $\overline{\boldsymbol{\Pi}}$) has to be updated based on the distributions of $\boldsymbol{\Theta}$ (and $\mathbf{B}$), they will be initialized manually in the first iteration.

4. EXPERIMENT RESULTS

4.1. Experiment Setup

In our experiment, we first choose partial images from the CAVE database [15]. In this database, images feature $L = 31$ spectral channel bands with size of 512×512. Thus, $M = N = 512$. Images selected for test are typical scenes in normal life and almost have reconstruction performance comparison in [8, 4, 9]. Then we get the low spatial resolution image $\mathbf{X}$ by averaging the original image over 8×8 spatially disjoint blocks and obtain the PAN image based on the Nikon D700 camera[1]. As the expectations of the dictionary $\mathbf{A}$ and the sparse representations are zero matrices, the low resolution image will be centered statistically and its mean will compensate for the final result to get the pan-sharpened image.

As for the model initialization, parameters $a_{\mathbf{X}}$, $b_{\mathbf{X}}$, $a_{\mathbf{N}}$, $b_{\mathbf{N}}$, $a_{\boldsymbol{\Theta}}$, $b_{\boldsymbol{\Theta}}$, $a_{\mathbf{B}}$, $b_{\mathbf{B}}$, c_{β}, d_{β}, c_{γ} and d_{γ} are all set to 10^{-6}.

[1]The data of spectral transform matrix $\mathbf{F}$ can be found at http://www.maxmax.com/spectral response.htm

Table 1. Algorithms performance comparison for partial images in CAVE database

Algorithms \ Image	Beers	Beans	Cloth	Faces	Hairs	Paints	Spools	Statue	Sushi	Toys
Proposed method	**1.8**	5.3	**3.5**	**1.1**	**1.3**	3.7	**4.4**	**1.0**	**2.0**	4.8
BSR.[4]	2.1	4.8	4.0	1.9	2.2	**3.2**	4.6	1.4	2.9	4.0
GSOMP.[9]	2.2	6.1	4.0	2.2	2.1	6.9	5.0	2.1	3.2	5.1
ADMM.[8]	-	**4.2**	9.5	3.4	2.3	4.5	5.3	4.3	-	**3.0**

The dimension dictionary of K and the iteration times T_{stop} are fixed to 20 and 100, respectively. Elements of $\overline{\mathbf{\Pi}}$ and $\overline{\mathbf{W}}$ are initialized to 0.5 while the dictionary is first set to the result of Principal Components Analysis on $\mathbf{X}$. We run our experiment on Matlab with Intel Core 3.6GHz i7 CPU and 8GB RAM. Partial codes for updating $\mathbf{\Theta}$ and $\mathbf{B}$ are written in C-MEX to enhance the looping efficiency.

Table 2. Average time consumption comparison

Algorithms	Proposed method	BSR.[4]
Time	786.27sec.(13min.)	10680sec.(178min.)

4.2. Performance with Comparison

Fig.1 offers the reconstruction error distribution of the image Chart & Toys in CAVE database [15]. In most areas, our reconstructed images have nearly no difference with the original images which are taken as ground truths. However, the value of absolute error increases along edges or in some delicate texture features, because pixels of these areas are largely blurred in low spatial resolution image and the spectral information is lost in the PAN image. Thus dictionary learning could not record accurate spectral features in these areas.

The performance results are shown in Table1 and the time consumption results are shown in Table 2 (The method proposed in [4] is called BSR. in abbreviation). We adopt root mean square error (RMSE) proposed in [8] to evaluate the pan-sharpening performance. From these two tables, we could find that our reconstruction performance is comparable in RMSE to other state-of-art algorithms while largely reducing the timing consumption by a factor of 13.58. In VBS algorithm [4], MCMC method is adopted, which requires long sampling times in each iteration (In [4], the sampling times are 128). However, our algorithm replaces the sampling process with single calculation of expectations. Although the matrix inverse brings in computationally cost, it could be simplified by Woodbury matrix identity [16]. Therefore, the speed of our method is largely enhanced. In some cases, the reconstruction performance of our algorithm is better than others because our dictionary is learned in consideration of the PAN image and the low spatial resolution image, which reserves spectral information from the low spatial resolution image and is more adaptive for the PAN image. In summary, our solution is promising for data fusion considering computational complexity and reconstruction performance.

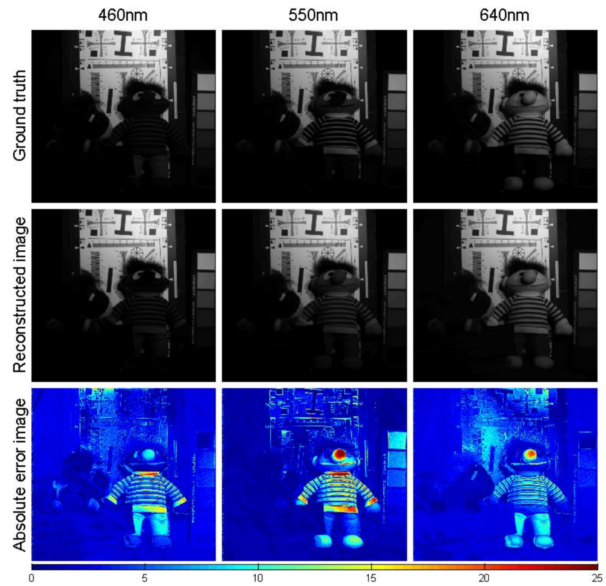

Fig. 1. Images reconstructed by our algorithm at 460, 550 and 640nm. Three rows of images are ground truths in CAVE database, results of our algorithm and the error distribution, respectively.

5. CONCLUSION

In this paper, we propose a Bayesian fusion model, which formulates the PAN image and the low spatial resolution HS image under the same Bayesian framework based on introduced "spike-and-slab" sparse prior and anisotropic Gaussian noise. To figure out the solution more efficiently, variational Bayesian algorithm is introduced to approximate the expectation of results. Experiments prove that the dictionary learning in consideration of two input images is more adaptive for image reconstruction and our scheme achieves comparable performance to other state-of-art algorithms while largely cutting down the computational complexity. For images in CAVE database, our algorithm can even run faster by up to 14 times with promising performance.

6. REFERENCES

[1] M. A. Bendoumi, M. He, and S. Mei, "Hyperspectral image resolution enhancement using high-resolution multispectral image based on spectral unmixing," *IEEE Trans. Geosci. and Remote Sens.*, vol. 52, no. 10, pp. 6574–6583, 2014.

[2] C. Chen, Y. Li, W. Liu, and J. Huang, "Image fusion with local spectral consistency and dynamic gradient sparsity," in *proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2014, pp. 2760–2765.

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

[4] N. Akhtar, F. Shafait, and A. Mian, "Bayesian sparse representation for hyperspectral image super-resolution," in *proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2015, pp. 3631–3640.

[5] 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. on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2011, pp. 2329–2336.

[6] M. A. Veganzones, M. Simoes, G. Licciardi, and et al., "Hyperspectral super-resolution of locally low rank images from complementary multisource data," in *proc. IEEE Int. Conf. on Image Processing (ICIP)*. IEEE, 2014, pp. 703–707.

[7] Q. Wei, N. Dobigeon, and JY. Tourneret, "Bayesian fusion of hyperspectral and multispectral images," in *proc. IEEE Int. Conf. on Acoust., Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 3176–3180.

[8] E. Wycoff, T. Chan, K. Jia, W. Ma, and Y. Ma, "A nonnegative sparse promoting algorithm for high resolution hyperspectral imaging," in *proc. IEEE Int. Conf. on Acoust., Speech and Signal Processing (ICASSP)*. IEEE, 2013, pp. 1409–1413.

[9] N. Akhtar, F. Shafait, and A. Mian, "Sparse spatio-spectral representation for hyperspectral image super-resolution," in *Computer Vision–ECCV 2014*, pp. 63–78. Springer, 2014.

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

[11] D. G Tzikas, A. C Likas, and N. P Galatsanos, "The variational approximation for bayesian inference," *IEEE Signal Processing Magazine*, vol. 25, no. 6, pp. 131–146, 2008.

[12] F. Scheipl, L. Fahrmeir, and T. Kneib, "Spike-and-slab priors for function selection in structured additive regression models," *Journal of the American Statistical Association*, vol. 107, no. 500, pp. 1518–1532, 2012.

[13] E. Million, "The hadamard product," *Course Notes*, vol. 3, 2007.

[14] S. Li, X. Tao, and J. Lu, "A low-complexity bayesian approach to large-scale sparse image reconstruction with structured constraints," in *proc. IEEE Global Comm. Conf. (GLOBECOM)*. IEEE, 2014, pp. 3115–3120.

[15] 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, 2010.

[16] G. H Golub and C. F Van Loan, *Matrix computations*, vol. 3, JHU Press, 2012.