IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 7, 2018, accepted March 14, 2018, date of publication March 19, 2018, date of current version April 18, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2817071

Hyperspectral and Multispectral Image Fusion Based on Low Rank Constrained Gaussian Mixture Model

BAIHONG LIN, XIAOMING TAO, (Member, IEEE), YIPING DUAN, AND JIANHUA LU, (Fellow, IEEE)
Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
Corresponding author: Xiaoming Tao (taoxm@mail.tsinghua.edu.cn)

This work was supported in part by the National Basic Research Project of China (973) under Grant 2013CB329006 and in part by the National Natural Science Foundation of China under Grant 61622110, Grant 61471220, and Grant 91538107.

ABSTRACT This paper attempts to fuse a multispectral image and an auxiliary hyperspectral image (HSI) with no requirement of image registration. Most previous studies solve this problem with sparsity-based methods. However, in this paper, a novel fusion framework is developed based on a Gaussian mixture model (GMM): First, the GMM is adopted to extract the spectral information from the input HSI. Low-rank constraints are imposed on the covariance matrices of the model to solve the computational problem in the expectation-maximization approach. Second, considering the spatial self-similarity, a structure-similarity regularization term is designed to further enhance the quality of the reconstructed image. To that end, a forward–backward splitting method is adopted to cut down the computational complexity of the optimization. The proposed method does not require two well-aligned images, thus, it will not be influenced by the registration errors between two fusing images. Experimental results of a simulated data set and an actual satellite (EO-1/Hyperion/ALI) data set show that the proposed method displays a stable performance and outperforms many state-of-the-art methods with acceptable computational complexity, when registration errors are taken into consideration.

INDEX TERMS
Hyperspectral image fusion, Gaussian mixture model, low rank constraint, local and nonlocal similarity, registration errors.

I. INTRODUCTION

Hyperspectral images (HSIs), with tens or hundreds of spectral channel bands, have been widely used in many computer vision tasks. Unfortunately, although these images have rich spectral information, spatial resolution is usually limited due to the technical trade-off between the resolutions of spectra and space in the imaging device. As a result, HSI super-resolution is proposed, working as a post-processing method to overcome the instrument limitation.

A popular way of HSI super-resolution (SR) is to fuse a low-spatial resolution HSI and a high-spatial resolution multispectral (or panchromatic) image (MSI) after image registration [1]. Most studies assume that two input images can be registered totally and perfectly. Under the assumption, the fusion issue turns into an ill-posed inverse problem, and a rational prior model on the target fusion image is the corresponding bottleneck [1]. So far, various prior models have been explored, including low rank constrained model [2]–[4], matrix factorization models [1], [5], [6], sparsity based models [7]–[10], total variation models [11], PCA-based models [12]–[14], etc. Apparently, if two input images can be registered totally and perfectly, the spatial resolution of the common region of these two image can be effectively enhanced by these fusion methods.

However, in practice, it is difficult to ensure the perfect registration between two fusing images, since this process can be easily affected by the differences in terms of angle of view, date of acquisition, and spectral coverage [15]. Therefore, the fusion accuracy of the above methods is inevitably compromised due to the registration errors [16]. Moreover, the overlapped scene between two images acquired by different sensors or satellites sometimes is very small, despite that these two images may share similar geographical feature.

In this case, the above registration-based fusion methods fail to work and cannot make full use of the nonlocal similarity laid in the input images [16], [17].

To overcome the above difficulties, various schemes have been proposed. Early studies promote to utilize component substitution fusion methods, since they are found relatively less sensitive to registration errors [18]. In the recent years, Zhang *et al.* [19] proposed a robust point-matching algorithm to reduce the registration errors. Chen *et al.* [16] further reduced the influence of registration errors by putting image registration and fusion in a unified optimization framework. Unfortunately, as the approach proposed by [16] was only available when the registration error was limited within a certain range, Zhang *et al.* [20] proposed a practical scheme for joint image registration and fusion later. To totally avoid registration errors, Akhtar *et al.* [21] and Huang *et al.* [22] proposed a novel framework without image registration requirements. In their approaches, the MSI is not directly fused with the HSI, instead, it is fused with a spectral dictionary learned from the low spatial resolution HSI. Obviously, how to obtain a good spectral dictionary is a key step for these methods. Thus, subsequently, Akhtar *et al.* [23], [24] proposed two Bayesian methods to further enhance the quality of the learned spectral dictionary. On the basis of spectral dictionary learning, to further enhance the fusion accuracy, Zhao *et al.* [15] introduced the joint regulation of spatial and spectral nonlocal similarities, whereas Fang *et al.* [25] proposed to obtain fusion results using super-pixels-based sparse representation. All the above schemes reduce the influence of registration errors in varying degrees. Particularly, the spectral dictionary methods [21]–[25] are less sensitive to registration errors, since the learned spectral dictionary has little or no relation to the translation and rotation of the input HSI. However, these kind of methods usually cost too much computational resource due to the optimization of the sparsity regularization. Furthermore, it remains open to question whether the learned spectral dictionary is the most efficient way to extract the spectral features from the input HSI.

In this paper, we develop a novel hyperspectral and multispectral image fusion framework based on a low rank constrained Gaussian mixture model (LR-GMM). The proposed framework has no registration requirement on the input HSI and the MSI. Thus, unlike the traditional registration-based methods [1]–[14], it will not be influenced by the registration errors between the HSI and the MSI. Moreover, different from the spectral dictionary approaches [21]–[25], the proposed method utilizes a novel efficient spectral model, the LR-GMM model, to extract spectral features from the low spatial resolution HSI, and jointly takes the local and nonlocal texture similarity of the MSI into consideration to further enhance the fusion accuracy. The main ideas of our work can be summarized as follows:

1) Gaussian mixture model (GMM) is adopted to efficiently extract the spectral feature from the HSI. Since the spectral pixels lay in a low dimensional subspace, we impose low rank constraints on the covariance matrices of GMM to cut down the complexity of the model.
2) To make full use of the spatial and spectral information of the MSI, the local and nonlocal structure similarity of spectral and spatial domains is incorporated into the proposed framework, working as a regularization term to improve the resolution of the fused image.
3) The proposed framework casts the HSI super-resolution as a quadratic optimization problem. A forward-backward splitting method is derived to cut down the computational complexity brought by the inverse of a very large matrix in the optimization.

The remainder of this paper is organized as follows. In Section II, we formulate the super-resolution problem discussed in our paper mathematically, and present the proposed approach and the optimization process. Experimental results and comparisons are given in Section III and the conclusion is drawn in Section IV.

II. PROPOSED LR-GMM BASED FUSION APPROACH

This section introduces the proposed LR-GMM based fusion framework. Define the input $m \times n \times L$ hyperspectral image (HSI) as $\mathbf{X} \in \mathbb{R}^{L \times mn}$, the input $M \times N \times l$ multispectral image (MSI) as $\mathbf{Y} \in \mathbb{R}^{l \times MN}$. $m \times n$ and $M \times N$ are the spatial sizes of two input images. Let $\mathbf{Z} \in \mathbb{R}^{L \times MN}$ denote the $M \times N \times L$ full resolution target image. Assuming that the noise brought by the measurement is additive and independent, and has a Gaussian distribution, then the observation model of $\mathbf{Z}$ can be formulated as:

$$\mathbf{Y} = \mathbf{FZ} + \mathbf{N}_y, \ \ \mathbf{X} = \mathbf{ZPQS} + \mathbf{N}_x \tag{1}$$

where $\mathbf{F}$ denotes the known relative spectral response between $\mathbf{Y}$ and $\mathbf{Z}$, $\mathbf{P}$ denotes a warping matrix, $\mathbf{Q}$ and $\mathbf{S}$ are an unknown image-blurring matrix and a down-sampling matrix respectively, $\mathbf{N}_x$ and $\mathbf{N}_y$ denote the isotropic Gaussian noise, i.e., $n_{y_{ij}} \sim \mathcal{N}(0, \lambda^{-1})$ and $n_{x_{ij}} \sim \mathcal{N}(0, \epsilon^{-1})$. In the problem discussed in our paper, $\mathbf{X}$ and $\mathbf{Y}$ are two not-well-aligned images, i.e., the warping matrix $\mathbf{P}$ is unknown and cannot be accurately obtained. This paper aims to obtain the full resolution image $\mathbf{Z}$ without having to estimate the warping matrix $\mathbf{P}$ and the blurring matrix $\mathbf{Q}$.

Apparently, the target full resolution image $\mathbf{Z}$ cannot be obtained simply resorting to Eqn.(1), since this is an ill-posed problem and requires prior information about the image $\mathbf{Z}$. According to Eqn.(1), although the spatial mapping between the images $\mathbf{X}$ and $\mathbf{Z}$ is unknown, $\mathbf{X}$ has similar spectral features with $\mathbf{Z}$, which have little or no relation to the mapping matrix $\mathbf{PQS}$. Moreover, since the target image $\mathbf{Z}$ is the high spectral-resolution version of the MS image $\mathbf{Y}$, the image $\mathbf{Y}$ shares the same geographical texture features with the image $\mathbf{Z}$. To make full use of the spectral information of the image $\mathbf{X}$ and the spatial information of the image $\mathbf{Y}$, we introduce two prior regularization terms for the target

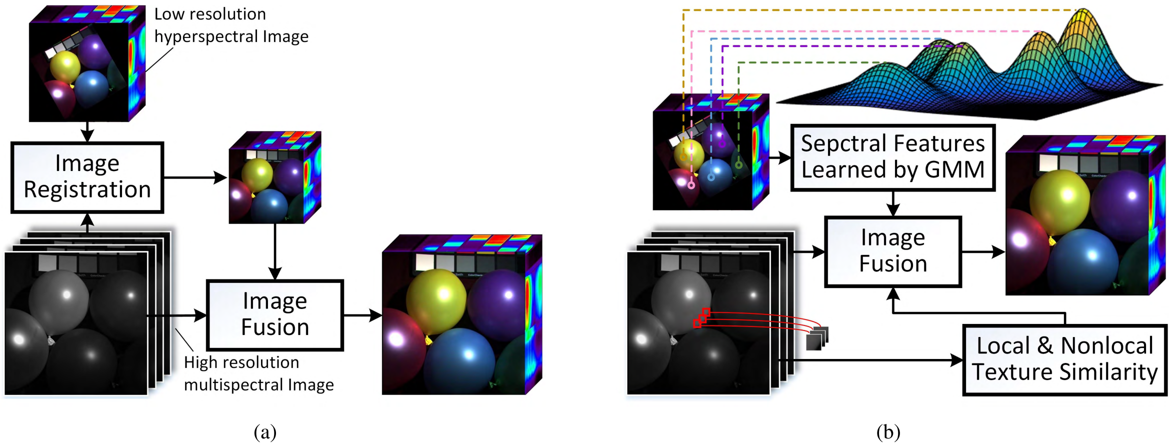


FIGURE 1. (a) The traditional registration-based fusion framework. (b) The proposed LR-GMM based fusion framework. The spectral features learned by Gaussian mixture model (GMM) has little or no relation to the image translation and rotation. Thus, the results are robust to registration errors.

image $\mathbf{Z}$, and formulate the prior of $\mathbf{Z}$ as:

$$\ln p(\mathbf{Z}) \propto -[g(\mathbf{Z}, \mathcal{S}(\mathbf{X})) + h(\mathbf{Z}, \mathcal{T}(\mathbf{Y}))] \tag{2}$$

where $g(\cdot)$ is designed to regularize the spectral features of $\mathbf{Z}$ while $h(\cdot)$ is introduced to regularize the texture features of $\mathbf{Z}$; $\mathcal{S}(\mathbf{X})$ is a designed operator to extract spectral information from the image $\mathbf{X}$ while $\mathcal{T}(\mathbf{Y})$ is a designed operator to extract local and nonlocal texture similarity from the image $\mathbf{Y}$. Since the relationship between the images $\mathbf{Y}$ and $\mathbf{Z}$ is known, according to Eqn.(1), the conditional distribution of $\mathbf{Y}$ given $\mathbf{Z}$ can be formulated as:

$$\ln p(\mathbf{Y}|\mathbf{Z}) \propto -\lambda\|\mathbf{Y} - \mathbf{F}\mathbf{Z}\|_F^2 \tag{3}$$

A maximum-a-posteriori(MAP)-based estimator is utilized to obtain the target image $\mathbf{Z}$. According to Bayes theory, $p(\mathbf{Z}|\mathbf{Y}) \propto p(\mathbf{Y}|\mathbf{Z})p(\mathbf{Z})$. Then, the fusion problem can be formulated as the following regularized least squares problem based on the introduced prior regularizations:

$$\min_{\mathbf{Z}} \lambda\|\mathbf{Y} - \mathbf{F}\mathbf{Z}\|_F^2 + g(\mathbf{Z}, \mathcal{S}(\mathbf{X})) + h(\mathbf{Z}, \mathcal{T}(\mathbf{Y})) \tag{4}$$

The differences between the registration-based fusion method and the proposed fusion framework are summarized in Fig.1. In the following passage, we will elaborate the introduced regularization terms, $g(\cdot)$ and $h(\cdot)$.

A. LEARNING SPECTRA VIA LOW RANK CONSTRAINED GMM

To extract the spectral features from the image $\mathbf{X}$, Gaussian mixture model (GMM) is utilized, i.e., each spectral pixel vector $\boldsymbol{x} \in \mathbb{R}^{L\times 1}$ is assumed to be drawn from a GMM with C mixture components,

$$p(\boldsymbol{x}) = \sum_{w=1}^{C} p(\boldsymbol{x}|w)p(w) = \sum_{w=1}^{C} \pi_w \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_w, \Sigma_w) \tag{5}$$

where $p(w) = \pi_w$ is the prior probability of the wth components, $p(\boldsymbol{x}|w) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_w, \Sigma_w)$ is a multivariate Gaussian distribution, $\boldsymbol{\mu}_w$ is the mean of the wth Gaussian component representing the spectral cluster center, Σ_w is the corresponding covariance matrix recording the spectral variance information. This modeling is reasonable, since each pixel in low-spatial resolution HSI usually contains more than one spectral component [26]. Furthermore, it has been proved that GMM, as a spectral clustering model, has similar qualities to the linear mixture model for characterization and classification of hyperspectral imagery [27]. Based on this modeling, the parameters of GMM can be easily learned by the EM algorithm, and $\mathcal{S}(\mathbf{X}) = \{\boldsymbol{\mu}_w, \Sigma_w\}_{w=1}^C$.

Unfortunately, in the EM algorithm, the inverse of covariance matrices will bring in a high computational cost, especially when the number of channel bands in the HSI is very large. Furthermore, since the spectral pixel vectors in the HSI tend to live in a low dimensional manifold [8], the inverse of covariance matrices will fail due to the rank deficiency. Hence, to avoid the above computational problems, we impose low rank constraints on the covariance matrices, i.e., $\Sigma_w = \Phi_w\Phi_w^T + \gamma\mathbf{I}_L$, where $\Phi_w \in \mathbb{R}^{L\times r}$ is a r-rank matrix, $\mathbf{I}_L$ is a $L \times L$ identity matrix, and γ is a small parameter set to 10^{-3}. Then, the parameters $\Theta = \{\pi_w, \boldsymbol{\mu}_w, \Phi_w\}_{w=1}^C$ can be efficiently learned by the modified EM approach based on the references [28]–[30]. (See the Appendix).

As the target image $\mathbf{Z}$ shares similar spectral features with $\mathbf{X}$, $\mathcal{S}(\mathbf{X}) = \{\boldsymbol{\mu}_w, \Phi_w\}_{w=1}^C$ can be adopted to regularize the image $\mathbf{Z} = [z_1, z_2, ..., z_{MN}]$ as the following form:

$$g(\mathbf{Z}, \mathcal{S}(\mathbf{X})) = \sum_{j=1}^{MN} \min_{w_j} (z_j - \boldsymbol{\mu}_{w_j})^T \Sigma_{w_j}^{-1} (z_j - \boldsymbol{\mu}_{w_j}) \tag{6}$$

Since z_j is unknown, finding the optimal value of w_j is a complicated optimization process. Fortunately, the observation $\mathbf{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_{MN}]$ is known and its conditional prior has a Gaussian distribution according to Eqn.(1). Thus, we can approximate w_j and simplify the regularization term (6) as a

Algorithm 1 LR-GMM-Based HSI Super-Resolution

Input: The images $\mathbf{X}$ and $\mathbf{Y}$, the spectral response $\mathbf{F}$

Learning $\Theta = \{\pi_w, \boldsymbol{\mu}_w, \Phi_w\}_{w=1}^C$ **from the image X:**

- Initialize Θ by the scheme in Section-III-A
- Update Θ iteratively using Eqn.(14), Eqn.(17) and Eqn.(16) shown in Appendix

Obtaining the weight matrix B from the image Y:

- Divide the image $\mathbf{Y}$ into overlapped patches and calculate the matrix $\mathbf{B}$ using Eqn.(8)

Solving the optimization problem in Eqn.(4):

- Initialize $\mathbf{Z}^{(0)}$ by solving Eqn.(10) with $\alpha = 0$
- Update $\mathbf{Z}^{(t)}$ iteratively using Eqn.(11)

Output: The high-spatial hyperspectral image $\mathbf{Z}$.

normal quadratic form by solving the following problem:

$$\widehat{w}_j = \arg\max_{w_j} \mathcal{N}(\mathbf{y}_j|\mathbf{F}\boldsymbol{\mu}_{w_j}, \mathbf{F}\Sigma_{w_j}\mathbf{F}^{\mathrm{T}} + \lambda^{-1}\mathbf{I}_l) \tag{7}$$

B. LOCAL AND NONLOCAL STRUCTURE SIMILARITY REGULARIZATION

Early studies have pointed out that spectral pixels are spatially correlated to their local and nonlocal similar neighbors in the high-spatial resolution HSI [15], [31]. These correlations can be regarded as a kind of priors on the texture features of the target image $\mathbf{Z}$. In our problem, as the input image $\mathbf{Y}$ is captured from the same scene of $\mathbf{Z}$, $\mathbf{Y}$ has the spatial self-similarity structure similar to $\mathbf{Z}$. Thus, we extract the structure similarity based on the MS image $\mathbf{Y}$ and utilize it as a kind of prior for the target image $\mathbf{Z}$.

To extract the self-similarity, the input $M \times N \times l$ MSI ($\mathbf{Y}$) is firstly divided into $p \times p \times l$ overlapped 3-D patches. Define the image patch centered at position i as $\mathbf{Y}_i \in \mathbb{R}^{l \times p^2}$. Then, for each $\mathbf{Y}_i$, we search its similar patches in a $S \times S \times l$ window, which form a set satisfying $\Delta_i = \{\mathbf{Y}_k | \|\mathbf{Y}_k - \mathbf{Y}_i\|_{\mathrm{F}}^2 < t_s p^2 l\}$., t_s is a given threshold. According to [15], $\mathbf{Y}_i$ can be predicted by the weighted average of its similar patches, i.e., $\sum_{\mathbf{Y}_k \in \Delta_i} b_{ik}\mathbf{Y}_k$, and its prediction error can be defined as the following, where $D_i = \sum_k b_{ik}$.

$$\left\|\mathbf{Y}_i - \sum_{\mathbf{Y}_k \in \Delta_i} b_{ik}\mathbf{Y}_k\right\|_{\mathrm{F}}^2, \quad b_{ik} = \frac{1}{D_i}\exp\left(-\frac{\|\mathbf{Y}_i - \mathbf{Y}_k\|_{\mathrm{F}}^2}{lp^2}\right) \tag{8}$$

Obviously, the prediction weights b_{ik} in Eqn.(8) can also be obtained in least-square way. However, we do not adopt this way for two reasons: First, as introduced in [15], the prediction error in Eqn.(8) should be small, but it cannot be set infinitely small because there is some trade-off among all the regularization terms in Eqn.(4). If the weights are obtained in least-square way, this prediction error will be too small, and the image will be too smooth and blurred. Second, the self-similarity extracted from the image $\mathbf{Y}$ should be shared by the target image $\mathbf{Z}$. The weights obtained in least-square way based on Eqn.(8) are effective for the image $\mathbf{Y}$, but they are not suitable for the image $\mathbf{Z}$, since the image $\mathbf{Y}$ is a projection of the image $\mathbf{Z}$ according to Eqn.(1). However, if the similarity between $\mathbf{Y}_k$ and $\mathbf{Y}_i$ is very high, the similarity between two corresponding image patches in $\mathbf{Z}$ is more likely to be very high. Considering the texture structure similarity shared by $\mathbf{Y}$ and $\mathbf{Z}$, we prefer to enhance the regularization weight of $\mathbf{Y}_k$ for $\mathbf{Y}_i$, if $\mathbf{Y}_k$ is more similar to $\mathbf{Y}_i$. In this case, it is better to choose to the exponential function to compute the weights.

As the self-similarity extracted from the image $\mathbf{Y}$ is shared by the image $\mathbf{Z}$, this error term can be extended to regularize the image $\mathbf{Z}$ based on Eqn.(1), i.e., the structure-similarity regularization term can be formulated as

$$h(\mathbf{Z}, \mathcal{T}(\mathbf{Y})) = \alpha\|\mathbf{Z}(\mathbf{I} - \mathbf{B})\|_{\mathrm{F}}^2 \tag{9}$$

where α is a preset positive regularization parameter, and $\mathbf{B} = [b_{ij}]$ is a sparse matrix. If $\mathbf{Y}_j \notin \Delta_i$, $b_{ij} = 0$. Otherwise, b_{ij} can be obtained based on Eqn.(8).

C. OPTIMIZATION AND COMPUTATIONAL COMPLEXITY

To solve the target image $\mathbf{Z}$ of the problem (4), an easy way is to set the corresponding gradient to zero, i.e.,

$$(\lambda\mathbf{F}^{\mathrm{T}}\mathbf{F} + \Sigma_{w_j}^{-1})z_j + \left[\alpha\mathbf{Z}\mathbf{G}\mathbf{G}^{\mathrm{T}}\right]_{.j} = \lambda\mathbf{F}^{\mathrm{T}}\mathbf{y}_j + \Sigma_{w_j}^{-1}\boldsymbol{\mu}_{w_j} \tag{10}$$

where $\mathbf{G} = \mathbf{I} - \mathbf{B}$, $[\cdot]_{.j}$ denotes the jth column vector of a matrix. If $\alpha = 0$ or $\mathbf{G}$ is an identity matrix, solving $\mathbf{Z}$ based on Eqn.(10) only brings in the order of computational complexity $\mathcal{O}(MNL^3)$. Unfortunately, $\mathbf{G}$ is obtained based on the input image $\mathbf{Y}$, and elements in $\mathbf{Z}$ are correlated by $\mathbf{G}$ and $\mathbf{F}$ in the above equation when $\alpha > 0$. Thus, the inverse problem of a very large matrix has to be encountered in this solution, which will bring in the unacceptable order of computational complexity $\mathcal{O}(M^3N^3L^3)$.

To avoid the large-matrix-inverse problem, we propose to use the forward-backward splitting (FBS) method instead of directly solving Eqn.(10). Let $f_1(\mathbf{Z}) = \lambda\|\mathbf{Y} - \mathbf{FZ}\|_{\mathrm{F}}^2 + g(\mathbf{Z}, \mathcal{S}(\mathbf{X}))$ and $f_2(\mathbf{Z}) = h(\mathbf{Z}, \mathcal{T}(\mathbf{Y}))$. Define four operators as $\mathcal{A} = \nabla f_1$, $\mathcal{B} = \nabla f_2$, $\mathcal{I}$ is an identity operator, and $\mathcal{J}_{\beta\mathcal{A}} = (\mathcal{I} + \beta\mathcal{A})^{-1}$. It can be easily proved that $\mathcal{A}$ is maximal monotone while $\mathcal{B}$ is a Lipschitz continuous monotone operator with Lipschitz constant $\|\nabla f_2\| = \|\alpha\mathbf{G}\mathbf{G}^{\mathrm{T}}\| > 0$. According to the FBS method [32], the iterative scheme, $\mathbf{Z}^{(n+1)} = \mathcal{J}_{\beta\mathcal{A}}(\mathcal{I} - \beta\mathcal{B})(\mathbf{Z}^{(n)})$, $\beta \in (0, 2/\|\nabla f_2\|)$, converges weakly to an element of the set of solutions $(\mathcal{A} + \mathcal{B})^{-1}(\{\mathbf{0}\})$, which can be equivalently written as:

$$\mathbf{Z}^{(n+1)} = \min_{\mathbf{Z}} \lambda\|\mathbf{Y} - \mathbf{FZ}\|_{\mathrm{F}}^2 + g(\mathbf{Z}, \mathcal{S}(\mathbf{X})) + \frac{1}{2\beta}\|\mathbf{Z} - \mathbf{Z}^{(n)} + \beta\nabla h(\mathbf{Z}^{(n)}, \mathcal{T}(\mathbf{Y}))\|_{\mathrm{F}}^2 \tag{11}$$

The problem in Eqn.(11) has a closed form solution, which brings in the order of computational complexity $\mathcal{O}(MNL^3)$. Let T_{opt} denote the number of iteration. $\mathbf{Z}^{(0)}$ is initialized based on Eqn.(10) with $\alpha = 0$. Then we can obtain the result with the order of computational complexity $\mathcal{O}(MNL^3T_{\mathrm{opt}})$

by iteratively updating $\mathbf{Z}^{(t)}$ using Eqn.(11). The optimization process is summarized in **Algorithm 1**.

III. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we introduce the experimental set up for the proposed method. Then, we compare our method with seven fusion methods, including the SASFM method [22], the GSOMP method [21], the BSR method [23], the SSR method [25], the HySure method [11], the BN method [33] and the BSDMF method [1]. The first four methods try to record spectral information based on dictionary learning. Similar to the proposed method, they have no registration requirement on the input HSI and MSI. For fairness, the sizes of dictionaries in these three methods are all set to 50. The last three comparison methods are registration-based fusion methods. In our experiment, to show the registration-error robustness, we compared the proposed method with these three registration-based fusion methods under different amounts of misregistration. All fusion methods are run on the MATLAB R2013a with Intel Core 3.6GHz i7 CPU and 16GB RAM.

A. IMPLEMENTATION DETAILS

Two problems arise when the proposed approach is put into implementation. First, as the EM algorithm is used to extract spectral features from $\mathbf{X}$, a good initialization is crucial to guarantee the effectiveness of the learned GMM model. In our experiment, as a single pixel is usually similar to its neighborhoods in the image, we propose to initialize the parameters of GMM based on the entropy rate super-pixels [34]. A super-pixel is defined as a cluster of pixels in a perceptually-uniform region in the image [34]. In our initialization, we generate $2C$ super-pixels of the input HSI by the approach in [34], and obtain the Euclidean distances for every two super-pixels' center locations. Then all super-pixels are partitioned into C clusters using hierarchical clustering based on the obtained Euclidean distances. As the number of spectral clusters in the images used in the following experiment is usually no more than 20, we set C to 20. Then, on the basis of the wth cluster, π_w and $\boldsymbol{\mu}_w$ are initialized by the corresponding empirical weight and mean respectively, while $\Phi_w \in \mathbb{R}^{L\times r}$ are obtained by truncated singular value decomposition (SVD) on the covariance matrix. Since pixels in a spectral cluster live in a low dimension manifold, the column size r of Φ_w should be very small. To reduce the computational cost brought by the matrix inverse in Eqn.(14), we set r to 1.

Second, there exist several model parameters, including the size of image patch p, the size of the search window S, the threshold t_s, and the regularization parameters λ and α. In Section-II-B, the parameters, p, S and t_s, are related to the texture structure regularization term. Theoretically, the size of the search window S should be set large so that more similar patches can be found all over the image. However, large S will bring in high computational cost. According to [15] and [31], considering the balance between performance and computational consumption, S is set to 18, and p is set to 3. t_s should be set small, and it is usually set to 10^{-3}.

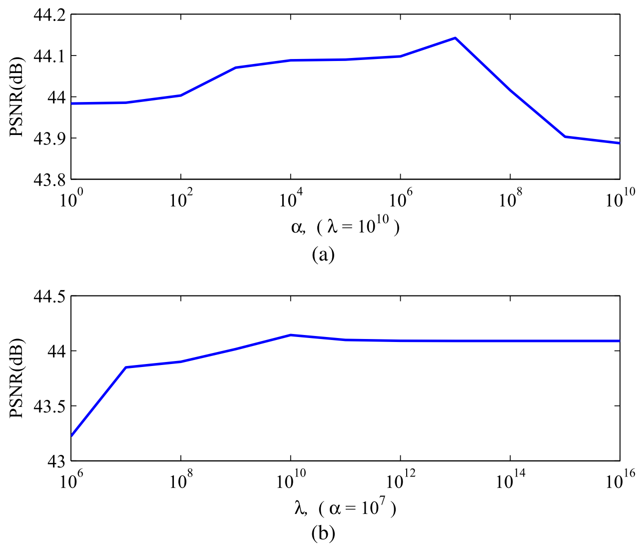


FIGURE 2. **The fusion results for different values of λ and α for the Balloon image in the CAVE dataset (Scale = 8). (a) PSNR for different values of α, λ is set to 10^{10}. (b) PSNR for different values of λ, α is set to 10^7.**

As for λ and α, on the basis of the above settings, we design a group of experiments to select the values of these two regularization parameters. In our experiment, we choose a real-life $512 \times 512 \times 31$ hyperspectral image as a ground truth and generate a low resolution HSI and a MSI according to the simulation method in Section-III-B. Then, these two generated images are fused using the proposed approach with different values of λ and α. The fusion results are shown in Fig.2. It can be found that the fusion results are relatively better when $\lambda = 10^{10}$ and $\alpha = 10^7$. Thus, in the following experiments, we set $\lambda = 10^{10}$ and $\alpha = 10^7$.

B. EXPERIMENTS WITH SIMULATED DATA

We test the performance of fusion methods using the CAVE database [35]. This database consists of 32 512-by-512 hyperspectral images formed by 31 spectral bands ranging from 400nm to 700nm at 10nm intervals in wavelength. It has been used to evaluate the fusion methods in [21]–[23].

In our experiment, each image in the CAVE database is taken as the ground truth. Low-spatial resolution HSIs and high-spatial resolution MSIs are generated according to Wald's protocol [36]. Specifically, for each image, three low-spatial resolution HSIs with 8, 16 and 32 spatial downsampling scales (denoted by s) are generated by 16×16, 32×32, and 64×64 Gaussian kernels with standard deviations of 3.40, 6.79 and 13.59, respectively. The high-spatial resolution MSI is generated by simulating the spectral response of the Nikon D700 camera.[1] To simulate the misregistration between the HSI and the MSI, we misalign these two input images along both x and y by d_p pixels at the scale of MSI, i.e., d_p/s at the scale of HSI. Then, three quality assessment metrics are

[1] Available at http://www.maxmax.com/spectral response.htm

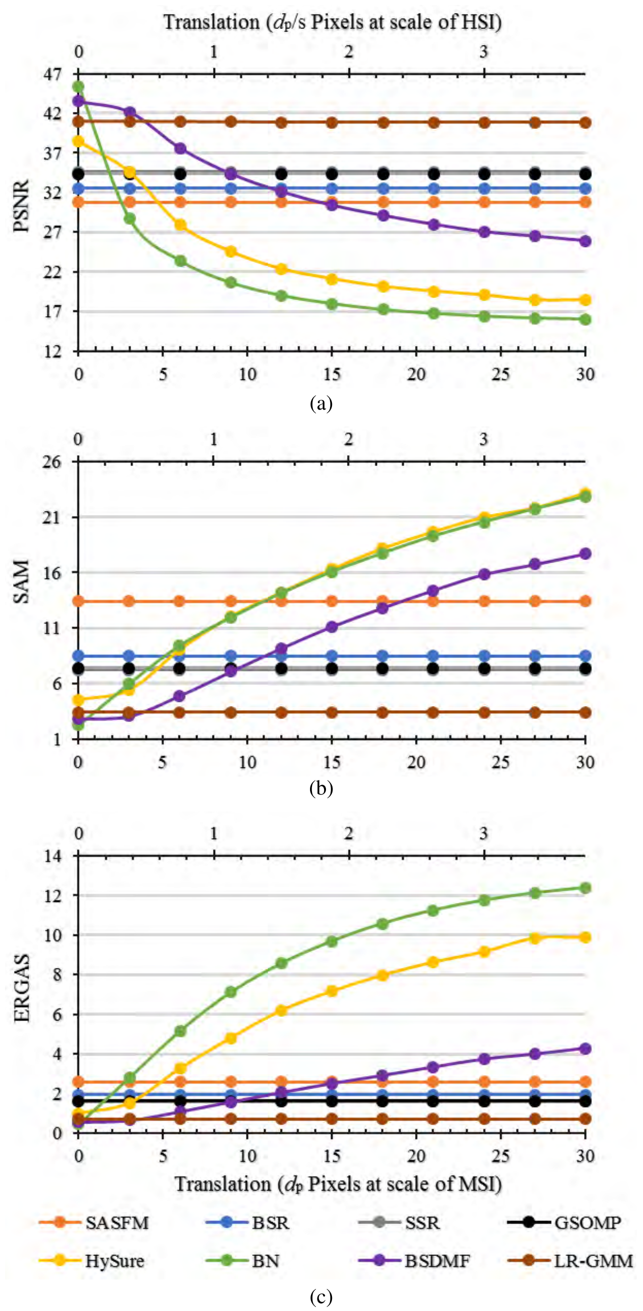


FIGURE 3. Quality Indexes for increasing amount of misregistration between the hyperspectral image and the RGB image. The results are based on the the Pompoms image in the CAVE database (Scale = 8). The bottom x-axis and the top x-axis represent the misregistration pixels at the scale of MSI and HSI, respectively.

TABLE 1. Evaluation of the results of the CAVE database*.

Scale	Method	Evaluation Index (Mean±Std.Dev)			
		PSNR(dB)	SAM(deg.)	ERGAS	Time
8	SASFM [22]	33.30±4.27	21.91±8.27	4.054±2.076	401
	GSOMP [21]	33.35±3.50	11.88±4.98	3.003±1.462	261
	BSR [23] ($Q = 64$)	34.86±2.99	10.28±3.07	2.616±1.537	9075
	SSR [25]	33.42±3.23	9.55±3.91	2.935±1.658	380
	HySure [11]	35.00±2.96	10.46±3.94	2.808±1.276	283
	BN [33]	29.35±3.04	7.53±2.85	4.987±1.935	**2**
	BSDMF [1]	**40.32±4.13**	**5.12±2.09**	**1.592±0.946**	**41**
	Proposed	**40.40±3.64**	**5.40±1.93**	**1.292±0.654**	79
16	SASFM [22]	32.74±3.83	21.03±7.55	1.931±0.841	515
	GSOMP [21]	32.92±4.78	12.48±4.70	1.713±1.520	279
	BSR [23] ($Q = 64$)	33.90±2.95	11.23±3.88	1.457±0.887	7490
	SSR [25]	33.74±3.67	10.48±4.05	1.396±0.718	233
	HySure [11]	35.27±3.17	15.44±6.11	1.377±0.657	300
	BN [33]	30.97±2.76	9.18±3.37	1.914±0.721	**2**
	BSDMF [1]	**38.22±4.64**	**6.93±3.51**	**1.006±0.713**	**45**
	Proposed	**39.80±3.92**	**5.43±1.90**	**0.687±0.370**	79
32	SASFM [22]	32.47±4.42	19.20±7.51	0.865±0.385	442
	GSOMP [21]	33.12±4.41	12.13±4.83	0.757±0.416	293
	BSR [23] ($Q = 64$)	33.60±3.36	12.08±6.46	0.755±0.529	5445
	SSR [25]	33.47±4.75	**9.40±3.09**	0.701±0.445	293
	HySure [11]	34.82±3.43	18.18±7.20	0.704±0.367	255
	BN [33]	32.82±2.80	10.98±3.78	0.718±0.308	**2**
	BSDMF [1]	**35.33±5.37**	9.93±5.19	**0.684±0.510**	**40**
	Proposed	**38.38±4.48**	**5.65±2.03**	**0.394±0.233**	79

*The above results are obtained with $d_p = Scale/2$. Specifically, the registration error between two input image is set to $Scale/2$ pixels at the scale of MSI, i.e., 0.5 pixels at the scale of HSI.

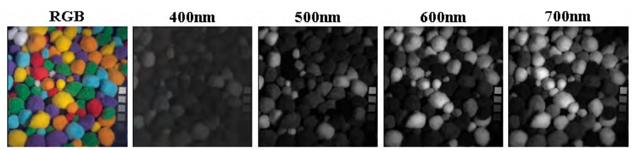


FIGURE 4. The Pompoms image in the CAVE database. The above images include the RGB image and the high-spatial resolution HSI at 400nm, 500nm, 600nm and 700nm.

adopted to evaluate the difference between the fused image and the ground truth, including PSNR, SAM and ERGAS, as used in [15]. The results are shown in Table 1, Fig.4 and Fig.5.

For simplicity, we show the means and standard deviations for the quality metrics of the CAVE's fusion results in Table 1. It can be found that the proposed method outperforms the other comparison methods in the quality of image reconstruction and costs acceptable computational time. Besides, since registration errors are taken into consideration in our simulation, most registration-based fusion methods perform badly and show relatively unstable results under different down-sampling scales.

To further show the robustness to registration errors, we display the results of the Pompoms image with different value of d_p in Fig.3. As the spatial mapping between the HSI and the MSI in the registration-based fusion framework is usually assumed to be known and can be made full use for image fusion, the registration-based methods usually outperform the proposed LR-GMM approach when two fusing images are perfectly aligned. Therefore, it can be found that the results of the BN and BSDMF methods in Fig.3 are better than the proposed when the pixels of translation are zeros (i.e., $d_p = 0$). However, the quality indexes of the registration-based methods tend to be worse with the increase of registration errors, whereas the proposed method remains stable in fusion accuracy and outperforms three registration-based methods when $d_p \geq 4$. These results imply that the

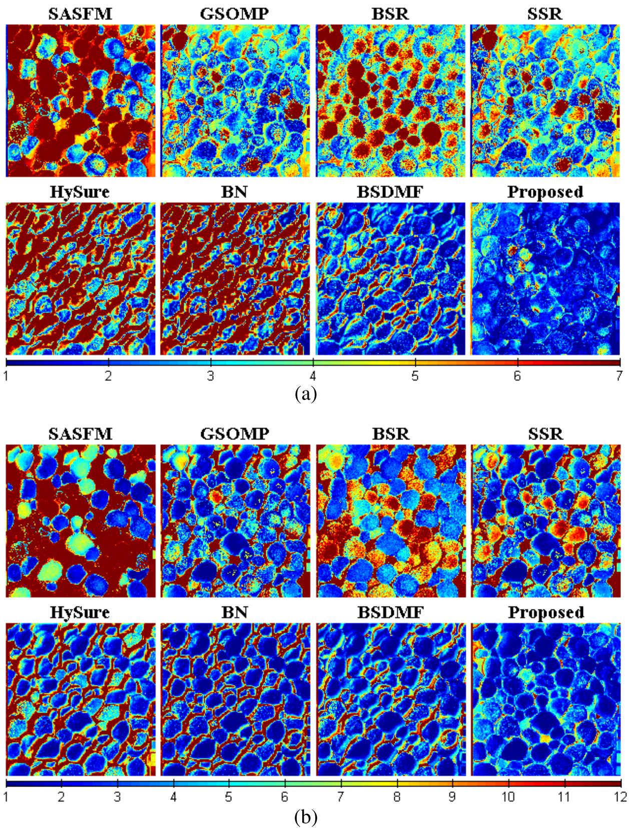


FIGURE 5. The results of the Pompoms image in the CAVE database (Scale = 8, $d_p = 4$). (a) RMSE images of different fusion methods. The RMSE values of all fusion methods are 8.59 (SASFM), 4.71 (GSOMP), 5.76 (BSR), 4.70 (SSR), 10.09 (HySure), 11.75 (BN), 3.21 (BSDMF), and 2.24 (LR-GMM). (b) SAM images of different fusion methods. The corresponding SAM values are 13.41 (SASFM), 7.47 (GSOMP), 8.49 (BSR), 7.42 (SSR), 8.99 (HySure), 7.26 (BN), 4.92 (BSDMF), and 2.97 (LR-GMM).

performance of the proposed method are better than those of the comparison methods when the input two images cannot be well aligned.

We also analyze the results of the comparison registration-based methods under different amounts of misregistration. The BN method works as the best one when the input images are perfectly aligned. These results are different from that in [1], since the sensor characteristics in our experiments are assumed to be accurately known and the measurement noise is not taken into consideration in the simulation. Unfortunately, as the BN method assumes two input images to be ideally aligned and requires an accurately-known spatial mapping between the HSI and the MSI, the performance of the BN method degrades sharply with the increase in the amount of misregistration in Fig.3, which implies that it has the poorest robustness to registration errors. The HySure method performs worse than the BSDMF method, since it also requires an accurately-known spatial mapping between the HSI and the MSI. However, compared with the BN method, the HySure method is relatively more robust to registration errors and has relatively good fusion results with clear texture due to the effectiveness of the total variation

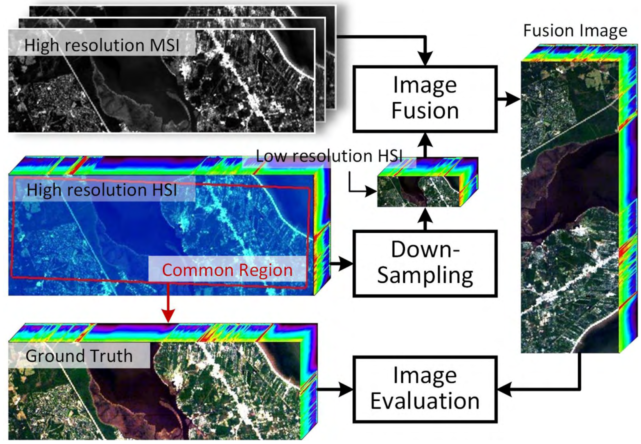


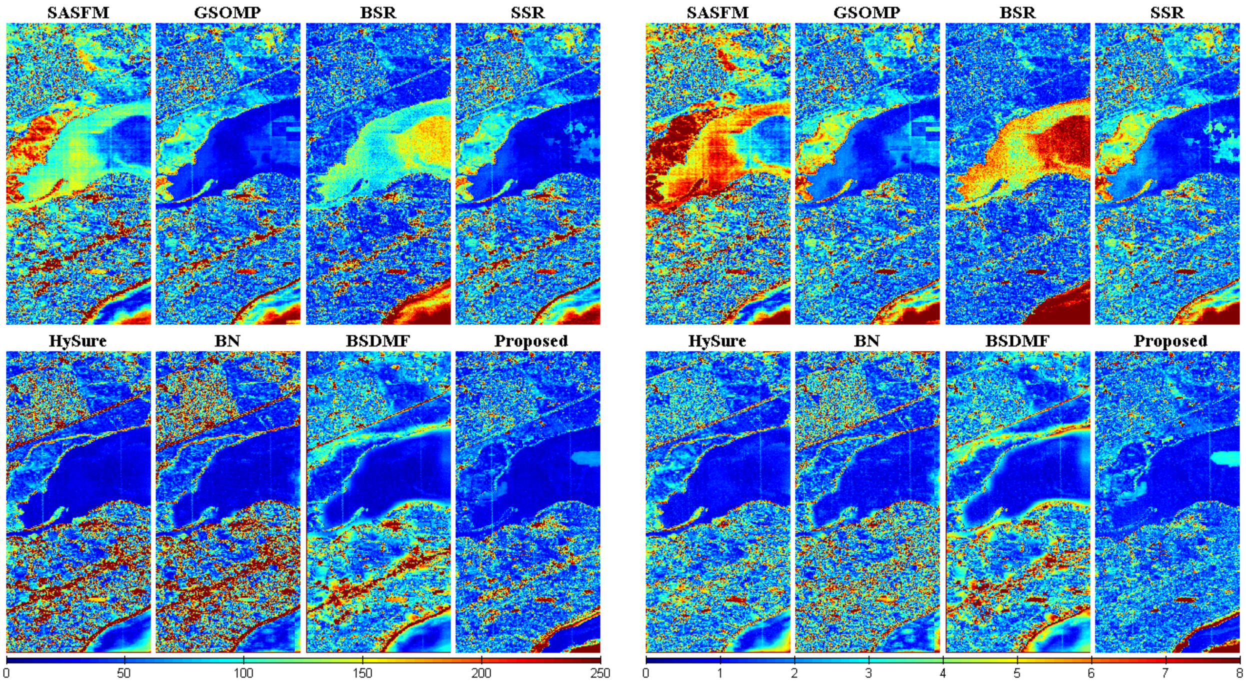
FIGURE 6. Evaluation process for the fusion methods on the actual satellite data.

regularization. The BSDMF methods is a matrix-factorization-based method with no requirement of the point spread function (PSF). Thus, it is not directly related to the estimation errors of the spatial mapping, and the corresponding results show great robustness to registration errors. It can be found that the BSDMF method is better than other comparison methods when $0 < d_p < 4$ in Fig.3. In Table 1, the BSDMF method has comparable results to those of the proposed method, when the down-sampling scale is set to 8. Therefore, it would be better to utilize the BSDMF method to fuse two slightly misaligned images. Nevertheless, the performance of the BSDMF method degrades with the increase of the down-sampling scale in Table 1 and tends to be worse than the proposed method when $d_p \geq 4$ in Fig.3. Apparently, it is not suitable for the cases when two input images have large scale difference or large registration errors.

C. EXPERIMENTS WITH ACTUAL SATELLITE DATA

In this section, we apply the proposed method to the fusion of EO-1/Hyperion and EO-1/ALI images.[2] The Hyperion image and the ALI image are all acquired over Bay St Louis (30°20′N, 89°20′W) at 16:15 pm on 15 October, 2001. The HSI captured by the Hyperion consists of 242 spectral bands ranging from 355.59nm to 2577.08nm in wavelength at 10nm interval, while the MSI captured by the ALI consists of 9 spectral bands [37]. Both of these two images are captured with a spatial resolution of 30 meters for all bands. According to [38], some of image bands have to be removed for practical applications due to their low image quality caused by water absorptions. In our experiment, since the image bands in the range of visible light are of high quality, for simplicity, we discuss the image fusion within the range of visible light. Specifically, we only keep the 8th-38th bands for the HSI corresponding to 427nm-732nm

[2]More information about EO-1/Hyperion and EO-1/ALI can be found at https://eo1.gsfc.nasa.gov/



(a) (b)

FIGURE 7. The results of the EO-1 satellite data set (Scale = 16) (a) the RMSE images displayed by rotating the results 90° clockwise. (b) the SAM images.

TABLE 2. Evaluation of the results of the EO-1 data set.

Scale	Method	Evaluation Index				
		PSNR	SAM	ERGAS	UIQI	Time
8	SASFM [22]	35.10	4.533	1.691	0.9607	84
	GSOMP [21]	38.13	2.653	1.075	0.9664	96
	BSR [23] ($Q = 128$)	37.28	2.873	1.149	**0.9677**	4928
	SSR [25]	37.96	2.854	1.112	0.9666	81
	HySure [11]	33.17	2.271	2.067	0.8642	153
	BN [33]	32.71	2.735	2.154	0.8597	**1**
	BSDMF [1]	**38.83**	**2.045**	**0.988**	0.9658	**17**
	Proposed	**39.03**	**2.056**	**1.004**	**0.9707**	23
16	SASFM [22]	32.04	4.492	0.880	0.9590	83
	GSOMP [21]	**34.64**	2.638	**0.566**	0.9574	87
	BSR [23] ($Q = 128$)	34.36	2.861	0.574	**0.9682**	3864
	SSR [25]	34.57	2.747	0.572	0.9572	54
	HySure [11]	30.33	**2.429**	1.024	0.8693	171
	BN [33]	29.87	2.864	1.065	0.8649	**1**
	BSDMF [1]	33.82	2.717	0.615	0.9467	**18**
	Proposed	**36.31**	**2.059**	**0.489**	**0.9714**	23
32	SASFM [22]	31.99	4.176	0.432	0.9568	77
	GSOMP [21]	34.12	2.389	0.296	0.9499	99
	BSR [23] ($Q = 128$)	34.17	3.044	**0.296**	**0.9662**	2861
	SSR [25]	**34.20**	**2.355**	0.297	0.9499	53
	HySure [11]	29.79	2.999	0.544	0.8681	188
	BN [33]	29.73	2.952	0.544	0.8672	**1**
	BSDMF [1]	32.05	3.497	0.378	0.9235	**19**
	Proposed	**35.93**	**2.107**	**0.258**	**0.9699**	36

in wavelength, and keep the 1th-3th bands for the MSI corresponding to 450nm-690nm in wavelength, according to [37]. The spatial sizes of the acquired high resolution HSI and MSI are set to 536 × 200 and 512 × 160, respectively. The images are shown in Fig.6. Obviously, these two images are misaligned and have different image sizes. The common region of these two images is set to 512 × 160, i.e., the real scene corresponding to the MSI is a partial scene corresponding to the HSI. Then, low-spatial resolution HSIs are obtained by directly down-sampling the acquired high resolution HSI at different scales, and the relative spectral response used for all fusion methods is estimated with the method proposed in [39]. Similar to the experiment in Section-III-B, to simulate the registration errors between the HSI and the MSI,, we misalign these two input images along y by 0.5 pixels at the scale of HSI, i.e, $s/2$ pixels at the scale of MSI.

We utilize the scheme shown in Fig.6 to evaluate the fusion accuracy. Several quality assessment metrics are adopted to quantify the fusion performance, including PSNR, SAM, ERGAS and UIQI, as used in [15] and [22]. The corresponding results are shown in Table 2 and Fig.7. Particularly, as the index of ERGAS is inversely proportional to the down-sampling scale [40], the corresponding values in Table 2 decrease with the increase of the scale. It can be found from the results that the reconstruction quality of the proposed method is not so outstanding as that of the simulation results in Section-III-B due to the unideal real imaging process and the estimation errors of the spectral response matrix. Nevertheless, the proposed method still outperforms other fusion methods based on the actual satellite data and tends to be better when the down-sampling scale is large.

IV. CONCLUSION

This paper proposes a novel hyperspectral image super-resolution (SR) framework by fusing a multispectral image (MSI) and an auxiliary hyperspectral image (HSI) without image registration. The proposed framework casts the SR problem into an optimization problem, in which a spectral regularization term is designed based on low rank constrained Gaussian mixture model (GMM) learned from the HSI while a texture regularization term is designed based on the local and nonlocal structure similarity in the MSI. Then, a forward-backward splitting method is adopted to cut down the computational complexity in the optimization. Exhaustive experiments show that the low rank constrained GMM is more efficient to extract the spectral information compared with many sparsity-based methods, and the proposed method outperforms other state-of-the-art methods in fusion quality and has acceptable computational cost, when the registration errors are taken into consideration.

APPENDIX

EM APPROACH FOR LOW RANK CONSTRAINED GMM]

The low rank constrained GMM can be equivalently written as the following form [28], [29],

$$p(\boldsymbol{x}) = \sum_{w=1}^{C} \pi_w \int \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_w + \Phi_w \boldsymbol{\epsilon}, \gamma \mathbf{I}) \mathcal{N}(\boldsymbol{\epsilon}|\mathbf{0}, \mathbf{I}) d\boldsymbol{\epsilon} \tag{12}$$

where $\boldsymbol{\epsilon} \in \mathbb{R}^{r\times 1}$ is an auxiliary latent variable, $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. To learn the parameters $\Theta = \{\pi_w, \boldsymbol{\mu}_w, \Phi_w\}_{w=1}^{C}$ from the image $\mathbf{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{mn}]$, we maximize the marginal log-likelihood based on Eqn(12), i.e.,

$$\widehat{\Theta} = \max_{\Theta} \sum_{i=1}^{mn} \ln \sum_{w_i=1}^{C} \int_{\boldsymbol{\epsilon}} p(w_i, \boldsymbol{\epsilon}_i, \boldsymbol{x}_i) d\boldsymbol{\epsilon} \tag{13}$$

The EM approach can be adopted to optimize the above maximization problem [29], [30]. In the expectation step (E-step), we approximate the posterior distributions of latent variables as follows:

$$\begin{aligned} p(w, \boldsymbol{\epsilon}|\boldsymbol{x}, \Theta^{(t-1)}) &= \rho_w \mathcal{N}(\boldsymbol{\epsilon}|\boldsymbol{\eta}_w, \boldsymbol{\Omega}_w) \\ \boldsymbol{\eta}_w &= (\gamma \mathbf{I} + \Phi_w^{\mathrm{T}} \Phi_w)^{-1} \Phi_w^{\mathrm{T}} (\boldsymbol{x} - \boldsymbol{\mu}_w) \\ \boldsymbol{\Omega}_w &= \mathbf{I} - (\gamma \mathbf{I} + \Phi_w^{\mathrm{T}} \Phi_w)^{-1} \Phi_w^{\mathrm{T}} \Phi_w \\ \rho_w &= \frac{\pi_w \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_w, \Phi_w \Phi_w^{\mathrm{T}} + \gamma \mathbf{I})}{\sum_{k=1}^{C} \pi_k \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \Phi_k \Phi_k^{\mathrm{T}} + \gamma \mathbf{I})} \end{aligned} \tag{14}$$

where $\{w_i, \boldsymbol{\epsilon}_i, \boldsymbol{x}_i\}$ are simplified as $\{w, \boldsymbol{\epsilon}, \boldsymbol{x}\}$. This step has the order of computational complexity $\mathcal{O}(mnLr^2)$. Then on the basis of approximation, calculate the expectation log-likelihood function as:

$$\mathcal{L}(\Theta|\Theta^{(t-1)}) = \sum_{i=1}^{mn} \mathbb{E}[\ln p(w_i, \boldsymbol{\epsilon}_i, \boldsymbol{x}_i)|w_i, \boldsymbol{\epsilon}_i; \Theta^{(t-1)}] \tag{15}$$

In the maximization step (M-step), we update the parameters $\Theta^{(t)}$ by finding the peak values of the expectation log-likelihood function, i.e., $\Theta^{(t)} = \arg\max_{\Theta} \mathcal{L}(\Theta|\Theta^{(t-1)})$. The calculation of the parameters $\Theta^{(t)}$ can be derived as

$$\begin{aligned} [\boldsymbol{\mu}_w^{(t)}, \boldsymbol{\Phi}_w^{(t)}] &= (\sum_{i=1}^{mn} \rho_{iw}^{(t)} \boldsymbol{x}_i \begin{bmatrix} 1 \\ \boldsymbol{\eta}_{iw}^{(t)} \end{bmatrix}^{\mathrm{T}}) \\ &\quad \times \left(\sum_{i=1}^{mn} \rho_{iw}^{(t)} \begin{bmatrix} 1 & (\boldsymbol{\eta}_{iw}^{(t)})^{\mathrm{T}} \\ (\boldsymbol{\eta}_{iw}^{(t)}) & \boldsymbol{\eta}_{iw}^{(t)} (\boldsymbol{\eta}_{iw}^{(t)})^{\mathrm{T}} + \boldsymbol{\Omega}_{iw}^{(t)} \end{bmatrix} \right)^{-1} \end{aligned} \tag{16}$$

$$\pi_w^{(t)} = \sum_{i=1}^{mn} \rho_{iw}^{(t)} \Bigg/ \sum_{w=1}^{C} \sum_{i=1}^{mn} \rho_{iw}^{(t)} \tag{17}$$

which has the order of computational complexity $\mathcal{O}(mnLr^2)$. Let T_{em} denote the number of iteration. Then the EM approach has the order of computational complexity $\mathcal{O}(mnLr^2 T_{\mathrm{em}})$.

REFERENCES

[1] B. Lin, X. Tao, M. Xu, L. Dong, and J. Lu, "Bayesian hyperspectral and multispectral image fusions via double matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5666–5678, Oct. 2017.

[2] 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.

[3] 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.

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

[5] 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.

[6] C. Lanaras, E. Baltsavias, and K. Schindler, "Hyperspectral super-resolution by coupled spectral unmixing," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3586–3594.

[7] 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.

[8] 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.

[9] 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.

[10] C. Zou and Y. Xia, "Bayesian dictionary learning for hyperspectral image super resolution in mixed Poisson–Gaussian noise," *Signal Process., Image Commun.*, vol. 60, pp. 29–41, Feb. 2018.

[11] M. Simoes, 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.

[12] 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./Oct. 2012, pp. 2133–2136.

[13] 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.

[14] M. González-Audícana, 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.

[15] Y. Zhao, J. Yang, and C.-W. Chan, "Hyperspectral imagery super-resolution by spatial–spectral joint nonlocal similarity," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2671–2679, Jun. 2014.

[16] 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.

[17] Y. Zhao, J. Yang, Q. Zhang, L. Song, Y. Cheng, and Q. Pan, "Hyperspectral imagery super-resolution by sparse representation and spectral regularization," *EURASIP J. Adv. Signal Process.*, vol. 2011, no. 1, pp. 1–10, 2011.

[18] S. Baronti, B. Aiazzi, M. Selva, A. Garzelli, and L. Alparone, "A theoretical analysis of the effects of aliasing and misregistration on pansharpened imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 446–453, Jun. 2011.

[19] K. Zhang, X. Li, and J. Zhang, "A robust point-matching algorithm for remote sensing image registration," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 2, pp. 469–473, Feb. 2014.

[20] Q. Zhang, Z. Cao, Z. Hu, Y. Jia, and X. Wu, "Joint image registration and fusion for panchromatic and multispectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 467–471, Mar. 2015.

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

[22] 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.

[23] 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.

[24] N. Akhtar, F. Shafait, and A. Mian, "Hierarchical Beta Process with Gaussian process prior for hyperspectral image super resolution," in *Computer Vision—ECCV*. Springer, 2016.

[25] L. Fang, H. Zhuo, and S. Li, "Super-resolution of hyperspectral image via superpixel-based sparse representation," *Neurocomputing*, vol. 273, pp. 171–177, Jan. 2018.

[26] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.

[27] S. G. Beaven, D. Stein, and L. E. Hoff, "Comparison of Gaussian mixture and linear mixture models for classification of hyperspectral data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, vol. 4. Jul. 2000, pp. 1597–1599.

[28] Z. Ghahramani and G. E. Hinton, "The EM algorithm for mixtures of factor analyzers," Dept. Comput. Sci., Univ. Toronto, Toronto, ON, Canada, Tech. Rep. CRG-TR-96-1, 1996.

[29] J. Yang *et al.*, "Compressive sensing by learning a Gaussian mixture model from measurements," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 106–119, Jan. 2015.

[30] M. J. Beal, "Variational algorithms for approximate Bayesian inference," M.S. thesis, Univ. College London, London, U.K., 2003.

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

[32] S. Setzer, "Splitting methods in image processing," Ph.D. dissertation, 2009.

[33] 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.

[34] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 2097–2104.

[35] 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.

[36] 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.

[37] J. A. Mendenhall and D. P. Ryan-Howard, "Earth observing-1 advanced land imager: Spectral response calibration," Lincoln Lab., Massachusetts Inst. Technol., Lexington, MA, USA, Project Rep. EO-1-2, 2000.

[38] R. Beck, "EO-1 user guide, version 2.3," Satellite Syst. Branch, USGS Earth Resour. Observ. Syst. Data Center, Sioux Falls, SD, USA, Tech. Rep., 2003.

[39] 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.

[40] 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 Remotely Sensed Images (SEE/URISCA)*, 2000, pp. 99–103.

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

XIAOMING TAO (M'09) received the B.S. degree from Xidian University, in 2003, and the Ph.D. degree from Tsinghua University, in 2008. From 2008 to 2009, she was a Researcher with France Télécom. From 2009 to 2011, she was a Post-Doctoral Research Fellow with the Department of Electrical Engineering, Tsinghua University. From 2012 to 2014, she was with Tsinghua University as an Assistant Professor, where she is currently an Associate Professor. Her research interests include wireless communication and networking, and multimedia signal processing.

YIPING DUAN received the B.S. degree from Henan Normal University, Xinxiang, China, in 2010, and the Ph.D. degree in computer science and technology from Xidian University and the Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education of China, Xian, China, in 2016. She is currently a Post-Doctoral Researcher with the Department of Electronic Engineering, Tsinghua University. Her research interests include semantic mining, machine learning, and SAR image processing.

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. Since 1989, he has been with the Department of Electronic Engineering, Tsinghua University, where he currently serves as a Professor. His current research interests include broadband wireless communication, multimedia signal processing, and wireless networking.

He is currently a Chief Scientist of the National Basic Research Program (973), China. He received the National Distinguished Young Scholar Fund from the NSF committee of China in 2005. He has published over 180 technical papers in international journals and conference proceedings. He is 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. He has been an active member of professional societies. He was one of the recipients of Best Paper Awards at the IEEE International Conference on Communications, Circuits and Systems in 2002, ChinaCom 2006, and the IEEE Embedded-Com 2012. He served as the Lead Chair for the General Symposium of IEEE ICC 2008 and a Program Committee Co-Chair for the 9th IEEE International Conference on Cognitive Informatics in 2010.