

Exploiting Low-Rank and Sparse Properties in Strided Convolution Matrix for Pansharpening

Feng Zhang, Haoran Zhang , Kai Zhang , Yinghui Xing , Jiande Sun , and Quanyuan Wu 

Abstract—Fusion of low spatial resolution multispectral (LR MS) and panchromatic (PAN) images to acquire high spatial resolution multispectral (HR MS) images has attracted increasing attention in recent years. In this article, we first utilize the form of convolution matrix (CM) to formulate the image fusion problem. In order to reduce the complexity of CM, the step size is introduced and strided convolution matrix (SCM) is constructed. Then, we explore the low-rank property in SCM and impose the prior on the spatial and spectral degradation model of LR MS and PAN images. Meanwhile, sparsity in SCM is considered to further enhance the local structures in the fused image. Finally, the proposed model is optimized efficiently by the alternative direction method of multipliers. By exploiting the low-rank and sparse priors in SCM of HR MS image, the local and global structures can be better preserved. The experimental results on the reduced-resolution and full-resolution datasets also show that the proposed method behaves well in qualitative and quantitative assessments.

Index Terms—Image fusion, low-rank and sparse priors, multispectral image, panchromatic (PAN) image, strided convolution matrix (SCM).

I. INTRODUCTION

MULTISOURCE images captured by different sensors can provide a more comprehensive understanding for the observed scene than the images from the single sensor. In order to efficiently use the redundant and complementary information among multisource images, image fusion technique is proposed, which can achieve efficient restoration [1], [2] and enhancement [3], [4] of different kinds of images. For example, image fusion is also applied to the integration of low spatial resolution multispectral (LR MS) images and panchromatic (PAN) images to generate high spatial resolution multispectral (HR MS) images.

Manuscript received November 16, 2020; revised January 10, 2021; accepted January 30, 2021. Date of publication February 9, 2021; date of current version March 3, 2021. This work was supported in part by the Natural Science Foundation of China under Grant 61901246 and Grant U1736122, in part by the China Postdoctoral Science Foundation under Grant 2019TQ0190 and Grant 2019M662432, in part by the Open Fund of National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, China, under Grant 20200208, and in part by the Natural Science Foundation for Distinguished Young Scholars of Shandong Province under Grant JQ201718. (Corresponding author: Kai Zhang.)

Feng Zhang, Haoran Zhang, Kai Zhang, and Jiande Sun are with the School of Information Science and Engineering, Shandong Normal University, Ji'nan 250358, China (e-mail: fengzhangpl@163.com; 1061701929@qq.com; zhangkainuc@163.com; jiandesun@hotmail.com).

Yinghui Xing is with the National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology and the School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: xyh_7491@nwpu.edu.cn).

Quanyuan Wu is with the College of Geography and Environment, Shandong Normal University, Ji'nan 250358, China (e-mail: wqy6420582@163.com).

Digital Object Identifier 10.1109/JSTARS.2021.3058158

So far, the fusion of LR MS and PAN images has been extensively studied and a great variety of methods are proposed. Generally, these methods can be divided into the following four categories [5]–[7].

- 1) Component substitution-based methods.
- 2) Multiresolution analysis (MRA)-based methods.
- 3) Degradation model-based methods.
- 4) Deep neural network (DNN)-based methods.

In the first category, the LR MS image is interpolated to match the size of the PAN image, and then, the spatial component of the LR MS image is synthesized by some methods [8]–[10]. For instance, adaptive Gram–Schmidt (GSA) [11] is proposed to adaptively estimate the combination weights of different bands. After a specific projection, the spatial component of the LR MS image is substituted by the PAN image and the desired HR MS image can be obtained by the corresponding inverse transform. The methods based on component substitution behave well in computational efficiency and implementation simplicity. But the spectral distortions usually appear in the fused results due to the differences in the spectral range between MS and PAN images [12].

For the second category of methods, the spatial information in the PAN image is extracted by some MRA tools and then injected into the interpolated LR MS image because it is assumed that only abundant spatial details are absent in LR MS image. Thus, the selection of spatial information analysis method has an important influence on the fused images. In [13], wavelet transform is considered to find proper spatial details, in which the spectral response of the MS image is utilized for weight calculation. Shah *et al.* [14] combined adaptive principal component analysis (APCA) and contourlet [15] to preserve the spatial details, which can overcome the limitation of the wavelet to better represent the directional information in images. In [16], Choi *et al.* also employed curvelet transform [17] to better represent the edges in the fused images because it is effective to improve the spatial resolution by enhancing the edges. Subsequently, some methods based on MRA [18]–[20] are also developed by estimating more accurate gain coefficients to achieve better fusion results. These methods have a better performance in spectral preservation but some spatial details from PAN images are excessively injected into LR MS images.

In the third category, LR MS and PAN images are viewed as the degradation results of the HR MS image in spatial and spectral domains, respectively. Therefore, HR MS images can be produced by solving the spatial and spectral degradation model, which is underdetermined and needs to impose valid priors for

the regularization of the solution space. In [21], the sparse prior is exploited and the reconstruction of HR MS images is realized by sharing the coding coefficients over HR/LR dictionary pair. Then, structure sparse prior is proposed in [22] to further explore the correlation among the bands of MS images. Besides, Wang *et al.* [23] used an autoregressive model to describe the local similarity in MS images. Yin [24] proposed a local adaptive sparse representation scheme to capture the similarity between LR MS and PAN images. Moreover, nonnegativity [25] and total variation [26] are also adopted to regularize the spatial and spectral degradation model. For instance, Chen *et al.* [27] proposed the dynamic gradient sparse prior to enhance the edge information in the fusion results. Deng *et al.* [28] further adopted the hyper-Laplacian prior to capture the differences between LR MS images and the fused images in gradient domain. For degradation model-based methods, the spatial and spectral information is preserved well in the fused images. However, the high computation complexity and generalization of priors limit their use.

Recently, the fourth category is arousing more attention for the fusion of LR MS and PAN images owing to the powerful representation ability of DNN. For example, Huang *et al.* [29] presented a stacked sparse denoising autoencoder in which DNN is constructed to learn the mapping between LR and HR image patches. Shao *et al.* [30] designed two DNNs to extract the salient features in LR MS and PAN images, and then, the learned residual is injected into the upsampled LR MS images. In [31], two separate branch networks are also introduced and the subpixel convolutional layer is added into DNN to obtain detailed feature maps. Subsequently, different networks [32]–[34] are further attempted for more subtle feature extraction. In [35], PanNet is constructed to model the spatial details in the fused images through residual learning. Then, PanNet is further improved in [36] by introducing the multiscale dilated blocks to capture the spatial information better. Moreover, Wei *et al.* [37] also employed residual learning to boost the performance of the fused images. He *et al.* [38] designed different convolution neural networks (CNNs) to extract proper details for injection. Deng *et al.* [39] also developed CNNs for pansharpening, where the difference images between PAN and LR MS images are fed into the networks. Guo *et al.* [40] combined the multiscale recursive blocks and the anisotropic total variation to overcome the spatial distortions. However, DNNs are usually trained on degraded data because there are no corresponding HR MS images for training in the real world. Thus, it is promising to propose new DNN architectures for real data training.

As a noticeable property in images, low-rank prior has been explored widely for image restoration [41]–[43] over the past decade. For example, Mao *et al.* [41] applied low-rank matrix completion on image and video recovery. The low-rank structure is also utilized for the fusion of LR MS and PAN images [44]–[46]. He *et al.* [44] simultaneously considered spatial and spectral sparse priors to acquire the fused image, in which the low-rank property is revealed by rearranging each band of the MS image into a matrix. However, the low-rank property cannot be satisfied well in [44] because MS images only contain several bands, e.g., 4 or 8 bands. In [46], a low-rank pansharpening

method is proposed from a new perspective of offset learning and spatial and spectral distortions are reduced. However, the formulations in these methods can only deal with the global structures in MS images, which ignore the local patterns in the images to be fused. Therefore, the fine details should be analyzed locally in order to improve the reconstruction quality of the fused image. Besides, the low-rank property in MS images can be achieved more decently by local modeling.

To cope with the mentioned issues, we propose a new LR MS and PAN image fusion method by exploring the low-rank and sparse properties in strided convolution matrix (SCM) in this article. In the proposed method, SCM is first developed to reduce the computational time for convolution matrix (CM) by introducing the step size, and then, the low-rank property in SCM constructed by MS images is deeply analyzed and verified. In the proposed method, the spatial and spectral degradation model is established from the perspective of the observation relationship among source images and HR MS images. Then, the low-rank prior is adopted to regularize the ill-posed model, which can deal with the global and local structures efficiently. Besides, a sparse representation model for the SCM of the fused image is introduced and combined with the basic fusion model, which can further improve the reconstruction performance to produce better fused images. Finally, an optimization algorithm is derived for the proposed model, which follows the framework of the alternative direction method of multipliers (ADMM) [47]. The fusion results on the datasets from different satellites demonstrate that the proposed method can achieve a better fusion performance. Some parameters are further investigated for a comprehensive understanding of the proposed method.

The remainder of the article is organized as follows. Section II gives the formulation of SCM in detail and analyzes the low-rank property of SCM constructed by MS images. The pansharpening method is proposed in Section III by taking the low-rank and sparse priors of SCM into account. Then, the optimization algorithm is designed to solve the proposed pansharpening model in Section IV. Experiments on different datasets are conducted for the comparison and analysis in Section V. Finally, conclusions are presented in Section VI.

II. STRIDED CONVOLUTION MATRIX

A. Convolution Matrix

CM is originally introduced in [48] to calculate a more accurate blur kernel for blind image deblurring. For an image $\mathbf{E} \in \mathbb{R}^{l_1 \times l_2}$ with size $l_1 \times l_2$, its convolution result with a kernel $f \in \mathbb{R}^{h_1 \times h_2}$ can be equivalently written as the following matrix multiplication form:

$$v(\mathbf{E} * f) = \mathbf{M}_h v(f) \quad (1)$$

where $v(\cdot)$ is the vectorization of matrix and $*$ stands for the convolution operation. \mathbf{M}_h is the corresponding matrix, which is called as the CM of \mathbf{E} . $h = \{h_1, h_2\}$ is the kernel size. Excluding the pixels influenced by boundary conditions, the size of \mathbf{M}_h is $(l_1 - h_1 + 1)(l_2 - h_2 + 1) \times h_1 h_2$. Besides, the first

convolution eigenvalue $\sigma_1(\mathbf{E})$ of the image \mathbf{E} is defined as

$$\sigma_1(\mathbf{E}) = \max_f \|\mathbf{E} * f\|_F, \text{ s.t. } \|f\|_F = 1 \quad (2)$$

where $\|\cdot\|_F$ is the Frobenius norm of a matrix. Then, the solution by maximizing the above equation is regarded as the first convolution eigenvector $f_1 \in \mathbb{R}^{h_1 \times h_2}$. Naturally, the i th ($i = 2, \dots, h_1 h_2$) convolution eigenvalue can be computed by

$$\sigma_i(\mathbf{E}) = \max_f \|\mathbf{E} * f_i\|_F, \text{ s.t. } \|f_i\|_F = 1, \langle f_i, f_j \rangle = 0 \forall j < i \quad (3)$$

where $\langle \cdot, \cdot \rangle$ is the inner product between two variables. In [48], it is proved that the convolution eigenvalues of the image \mathbf{E} are exactly the singular values of its associated CM \mathbf{M}_h . So, the global low-rank property in \mathbf{E} can be locally depicted by minimizing the nuclear norm of \mathbf{M}_h , because CM is composed by aggregating local patches together. Then, the nuclear norm minimization (NNM) of CM is used for image and video completion and produces satisfactory results [49]. However, the computational complexity for the optimization of NNM cannot be neglected, because CM will be very large by the formulation in (1), especially for high-dimensional data or large kernel.

B. Strided Convolution Matrix

In order to reduce the complexity, the sampling distance is increased and the convolution of f with image \mathbf{E} is implemented with step size $s = \{s_1, s_2\}$ ($1 \leq s_1 \leq h_1, 1 \leq s_2 \leq h_2$), which can be reformulated as

$$v(\mathbf{E}_s * f) = \mathbf{M}_{h,s} v(f) \quad (4)$$

where \mathbf{E}_s denotes the sampled elements in \mathbf{E} for the corresponding convolution with f , which is controlled by the step size s . s_1 and s_2 are the step size on horizontal and vertical directions, respectively. Then, $\mathbf{M}_{h,s}$ is the SCM, whose size is $((l_1 - h_1)/s_1 + 1)((l_2 - h_2)/s_2 + 1) \times h_1 h_2$. Obviously, CM is a special case of SCM when s_1 and s_2 are both set as 1. When the step size is larger than 1, $\mathbf{E}_s * f$ only contains part of the information in $\mathbf{E} * f$. However, the entire information of \mathbf{E} can be found in SCM. Thus, the low-rank property in \mathbf{E} also can be preserved in SCM.

In order to validate the low-rank prior in SCM, 500 MS images with size $256 \times 256 \times 4$, collected from Sundarbans, India, by QuickBird satellite, are employed and converted into their corresponding SCM, in which h_1 and h_2 are set as 16 and the step sizes s_1 and s_2 increase from 1 to 16. Then, singular value decomposition (SVD) is implemented on SCMs with the same s and the means of singular values from the same index are calculated and shown in Fig. 1. From Fig. 1(a), we can see that the singular values dramatically decrease with the increase of index, particularly for the indexes smaller than 15. Besides, we also display the ratio of the other singular values to the first singular value in Fig. 1(b). It can be found that the values are very close to zero when the index is larger than 50, which implies the low-rank property in SCM.

To ensure the low-rank property in CM and SCM, singular value thresholding (SVT) [50] is utilized to update them in most of the optimization algorithms, which

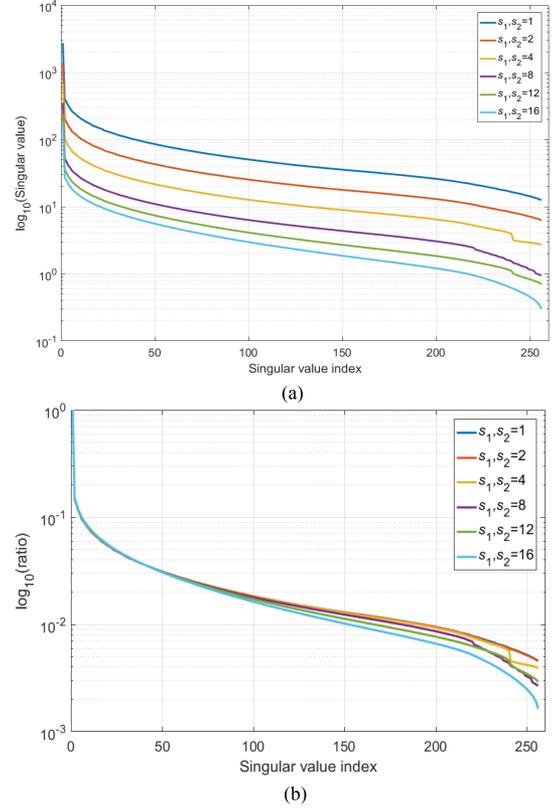


Fig. 1. Low-rank property in SCM with different s .

usually dominate their complexity. For CM, the complexity is $o(2(l_1 - h_1 + 1)(l_2 - h_2 + 1)h_1^2 h_2^2)$ in SVT step. The update of SCM through SVT has a complexity of $o(2(l_1 - h_1 + s_1)(l_2 - h_2 + s_2)h_1^2 h_2^2 / s_1 s_2)$. Besides, CM will need larger memory compared with SCM [50]. Thus, the memory requirements and complexity will be reduced a lot by using SCM with the increasing of s .

III. IMAGE FUSION USING LOW-RANK AND SPARSE PROPERTIES IN SCM

In this section, the low-rank property in SCM is introduced for the fusion of LR MS and PAN images. Besides, the sparse representation model is also considered for the local spatial and spectral structure preservation. Finally, the fusion method for LR MS and PAN images is proposed by combining the above degradation model and priors in SCM.

A. Low-Rank SCM for Image Fusion

In this article, LR MS and PAN images are denoted as $\mathcal{L} \in \mathbb{R}^{m \times n \times 4}$ and $\mathcal{P} \in \mathbb{R}^{M \times N}$, respectively. Then, the size is $M \times N \times 4$ for the desired HR MS image \mathcal{H} . Here, $M = rm$ and $N = rn$ are the width and height of the fused image. r is the spatial resolution ratio between LR MS and HR MS images. In the proposed method, \mathcal{P} is rearranged into SCM with the kernel size $h = \{h_1, h_2\}$ and step size $s = \{s_1, s_2\}$, which is denoted by $\mathbf{P}_{h,s} \in \mathbb{R}^{K \times h_1 h_2}$. K equals

$((M - h_1)/s_1 + 1)((N - h_2)/s_2 + 1)$. Taking the spatial resolution difference into consideration, the SCM of \mathcal{L} is $\mathbf{L}_{h^L, s^L} \in \mathbb{R}^{4K \times h_1^L h_2^L}$, in which $h^L = \{h_1^L = h_1/r, h_2^L = h_2/r\}$ and $s^L = \{s_1^L = s_1/r, s_2^L = s_2/r\}$. For convenience, the subscripts of the symbol for SCM are omitted and LR MS and PAN images are represented as \mathbf{L} and \mathbf{P} .

For LR MS and PAN images, the spatial and spectral degradation relationships can be defined as

$$\mathbf{L} = \mathbf{HBR} + n_1 \quad (5)$$

$$\mathbf{P} = \mathbf{SH} + n_2 \quad (6)$$

where $\mathbf{H} \in \mathbb{R}^{4K \times h_1 h_2}$ is the SCM of the desired HR MS image. $\mathbf{R} \in \mathbb{R}^{h_1 h_2 \times h_1^L h_2^L}$ and $\mathbf{B} \in \mathbb{R}^{h_1 h_2 \times h_1 h_2}$ are the downsampling matrix and blur matrix. The spectral response matrix is $\mathbf{S} \in \mathbb{R}^{K \times 4K}$. n_1 and n_2 indicate the Gaussian white noises.

According to the analysis in Fig. 1, it is assumed that \mathbf{H} possesses the low-rank property in the proposed method. Then, the basic fusion model can be written as follows by combining the low-rank constraint:

$$\min_{\mathbf{H}} \text{rank}(\mathbf{H}) + \frac{\alpha}{2} \|\mathbf{L} - \mathbf{HBR}\|_F^2 + \frac{\beta}{2} \|\mathbf{P} - \mathbf{SH}\|_F^2 \quad (7)$$

where $\text{rank}(\mathbf{H})$ is the rank of \mathbf{H} . α and β are the tradeoff parameters. Obviously, (7) is difficult to optimize due to its nonconvexity. Then, the rank constraint is relaxed by minimizing the nuclear norm under broad conditions, which is rewritten as

$$\min_{\mathbf{H}} \|\mathbf{H}\|_* + \frac{\alpha}{2} \|\mathbf{L} - \mathbf{HBR}\|_F^2 + \frac{\beta}{2} \|\mathbf{P} - \mathbf{SH}\|_F^2 \quad (8)$$

where $\|\cdot\|_*$ denotes the nuclear norm which is the sum of all singular values of a matrix. In (8), the global and local structures in the fused HR MS image can be simultaneously exploited by the low-rank constraint on \mathbf{H} . To obtain more accurate solutions of (8), more priors, such as sparsity and similarity, should be considered.

B. Sparse Prior

In the proposed method, sparse representation is utilized. It is proved that sparse prior can achieve a great performance in the ill-posed problem of image restoration [51], [52]. In the proposed method, one row in \mathbf{H} is supposed to have a sparse representation over a dictionary \mathbf{D} with the proper size, which is defined as

$$\min_{\mathbf{a}_i} \|\mathbf{a}_i\|_0 \text{ s.t. } \mathbf{h}_i^T = \mathbf{D}\mathbf{a}_i + n_3 \quad (9)$$

where \mathbf{h}_i is the i th row in \mathbf{H} . T stands for the transpose operation. \mathbf{a}_i is the corresponding sparse coefficient. The L_0 -norm of a vector is expressed by $\|\cdot\|_0$. n_3 is the noise. Accounting all rows of \mathbf{H} , (9) can be equivalently rewritten as follows with the relaxation of sparse constraint:

$$\min_{\mathbf{a}_i} \sum_{i=1}^{4K} \|\mathbf{a}_i\|_1 \text{ s.t. } \mathbf{H}^T = \mathbf{D}\mathbf{A} + n_3 \quad (10)$$

where $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \dots, \mathbf{a}_N] \in \mathbb{R}^{L \times 4K}$. L is the length of sparse coefficient, which also denotes the total number of atoms in \mathbf{D} . L_1 -norm $\|\cdot\|_1$ is the sum of the absolute values of all

vector entries. In the proposed method, \mathbf{D} is pretrained by using \mathbf{P}^T . Because the differences between PAN and the fused images are mainly caused by spectral range, they share the same spatial information and can be represented efficiently by the same dictionary. Therefore, \mathbf{D} is learned from PAN image by online dictionary learning [53] in advance.

C. Proposed Model

By incorporating the above priors, the well-posed LR MS and PAN image fusion model is modeled as

$$\begin{aligned} \min_{\mathbf{H}, \mathbf{A}} \|\mathbf{H}\|_* + \frac{\alpha}{2} \|\mathbf{L} - \mathbf{HBR}\|_F^2 + \frac{\beta}{2} \|\mathbf{P} - \mathbf{SH}\|_F^2 \\ + \lambda \sum_{i=1}^{4K} \|\mathbf{a}_i\|_1 \\ \text{s.t. } \mathbf{H}^T = \mathbf{D}\mathbf{A} + n_3 \end{aligned} \quad (11)$$

where λ controls the sparse regularization. According to the formulation in (11), spatial and spectral fidelity terms are regularized by the low-rank property in SCM and the sparse representation model for \mathbf{H} is adopted to ensure more reliable recovery.

IV. OPTIMIZATION ALGORITHM

In this section, the optimization algorithm is derived for the proposed fusion model. Because the problem in (11) is highly nonconvex, it is solved alternatively and iteratively through ADMM. Considering the coupling between fidelity terms and low-rank constraint, an auxiliary variable \mathbf{X} is introduced to replace \mathbf{H} in the first term of (11). Then, (11) is reformulated as

$$\begin{aligned} \min_{\mathbf{H}, \mathbf{A}, \mathbf{X}} \|\mathbf{X}\|_* + \frac{\alpha}{2} \|\mathbf{L} - \mathbf{HBR}\|_F^2 + \frac{\beta}{2} \|\mathbf{P} - \mathbf{SH}\|_F^2 \\ + \lambda \sum_{i=1}^{4K} \|\mathbf{a}_i\|_1 \\ \text{s.t. } \mathbf{H}^T = \mathbf{D}\mathbf{A} + n_3, \mathbf{H} = \mathbf{X}. \end{aligned} \quad (12)$$

Then, the augmented Lagrange function is

$$\begin{aligned} \min_{\mathbf{H}, \mathbf{A}, \mathbf{X}, \mathbf{Y}_1, \mathbf{Y}_2} \|\mathbf{X}\|_* + \frac{\alpha}{2} \|\mathbf{L} - \mathbf{HBR}\|_F^2 + \frac{\beta}{2} \|\mathbf{P} - \mathbf{SH}\|_F^2 \\ + \lambda \sum_{i=1}^{4K} \|\mathbf{a}_i\|_1 + \frac{\mu_1}{2} \|\mathbf{H}^T - \mathbf{D}\mathbf{A}\|_F^2 + \langle \mathbf{Y}_1, \mathbf{H}^T - \mathbf{D}\mathbf{A} \rangle \\ + \frac{\mu_2}{2} \|\mathbf{H} - \mathbf{X}\|_F^2 + \langle \mathbf{Y}_2, \mathbf{H} - \mathbf{X} \rangle \end{aligned} \quad (13)$$

where \mathbf{Y}_1 and \mathbf{Y}_2 are Lagrange multipliers. μ_1 and μ_2 are the penalty parameters. According to the framework of ADMM, these variables are updated by fixing the other variables.

For \mathbf{H} , its subproblem can be simplified as

$$\begin{aligned} F_{\mathbf{H}} = \frac{\alpha}{2} \|\mathbf{L} - \mathbf{HBR}\|_F^2 + \frac{\beta}{2} \|\mathbf{P} - \mathbf{SH}\|_F^2 \\ + \frac{\mu_1}{2} \|\mathbf{H}^T - \mathbf{D}\mathbf{A}\|_F^2 + \langle \mathbf{Y}_1, \mathbf{H}^T - \mathbf{D}\mathbf{A} \rangle \\ + \frac{\mu_2}{2} \|\mathbf{H} - \mathbf{X}\|_F^2 + \langle \mathbf{Y}_2, \mathbf{H} - \mathbf{X} \rangle. \end{aligned} \quad (14)$$

Then, the derivative of $F_{\mathbf{H}}$ w.r.t. \mathbf{H} is

$$\frac{\partial F_{\mathbf{H}}}{\partial \mathbf{H}} = \alpha \mathbf{H} \mathbf{B} \mathbf{R} \mathbf{R}^T \mathbf{B}^T - \alpha \mathbf{L} \mathbf{R}^T \mathbf{B}^T + \beta \mathbf{S}^T \mathbf{S} \mathbf{H} - \beta \mathbf{S}^T \mathbf{P} + \mu_1 \mathbf{H} - \mu_1 \mathbf{A}^T \mathbf{D}^T + \mathbf{Y}_1^T + \mu_2 \mathbf{H} - \mu_2 \mathbf{X} + \mathbf{Y}_2. \quad (15)$$

By making (15) equal to zero, the optimal \mathbf{H} can be calculated by

$$\mathbf{Z}_1 \mathbf{H} + \mathbf{H} \mathbf{Z}_2 = \mathbf{Z}_3 \quad (16)$$

where

$$\mathbf{Z}_1 = \beta \mathbf{S}^T \mathbf{S} + \mu_1 \mathbf{I} \quad (17)$$

$$\mathbf{Z}_2 = \alpha \mathbf{B} \mathbf{R} \mathbf{R}^T \mathbf{B}^T + \mu_2 \mathbf{I} \quad (18)$$

$$\mathbf{Z}_3 = \alpha \mathbf{L} \mathbf{R}^T \mathbf{B}^T + \beta \mathbf{S}^T \mathbf{P} + \mu_1 \mathbf{A}^T \mathbf{D}^T + \mu_2 \mathbf{X} - \mathbf{Y}_1^T - \mathbf{Y}_2 \quad (19)$$

where \mathbf{I} is the indent matrix with matched size. Obviously, (16) is typical of the Sylvester equation whose sufficient and necessary condition for a unique solution is that \mathbf{Z}_1 and $-\mathbf{Z}_2$ do not have the same eigenvalues [54]. Fortunately, \mathbf{Z}_1 and \mathbf{Z}_2 are both positive definite matrices. Thus, (16) has a unique solution for \mathbf{H} .

The subproblem of \mathbf{X} can be rewritten as

$$F_{\mathbf{X}} = \|\mathbf{X}\|_* + \frac{\mu_2}{2} \|\mathbf{H} - \mathbf{X}\|_F^2 + \langle \mathbf{Y}_2, \mathbf{H} - \mathbf{X} \rangle. \quad (20)$$

\mathbf{X} can be updated directly by SVT in [50], which is

$$\mathbf{X} = \mathbf{U} \mathcal{S}_{\tau}(\boldsymbol{\Sigma}) \mathbf{V}^T \quad (21)$$

where the SVD of $\mathbf{Z}_{\mathbf{X}}$ is $\mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^T$ and $\mathbf{Z}_{\mathbf{X}} = \mathbf{H} + \frac{\mathbf{Y}_2}{\mu_2}$. $\tau = \frac{1}{\mu_2}$. The soft-thresholding operator is defined as

$$\mathcal{S}_{\tau}(x) = \begin{cases} x - \tau, & \text{if } x > \tau \\ x + \tau, & \text{if } x < -\tau \\ 0, & \text{otherwise.} \end{cases} \quad (22)$$

For \mathbf{a}_i , the subproblem is

$$F_{\mathbf{a}_i} = \lambda \|\mathbf{a}_i\|_1 + \frac{\mu_1}{2} \|\mathbf{w}_i - \mathbf{D} \mathbf{a}_i\|_F^2 \quad (23)$$

where \mathbf{w}_i is the i th column in \mathbf{W} and $\mathbf{W} = \mathbf{H}^T + \frac{\mathbf{Y}_1}{\mu_1}$. Equation (23) can be quickly solved by the least angle regression (LARS) algorithm [55] with accurate reconstruction.

Besides, the multipliers \mathbf{Y}_1 and \mathbf{Y}_2 are updated by

$$\mathbf{Y}_1 = \mathbf{Y}_1 + \mu_1 (\mathbf{H}^T - \mathbf{D} \mathbf{A}) \quad (24)$$

$$\mathbf{Y}_2 = \mathbf{Y}_2 + \mu_2 (\mathbf{H} - \mathbf{X}). \quad (25)$$

Then, the penalty parameters increase with a small gain

$$\mu_1 = \min(\rho \mu_1, \mu_{\max}) \quad (26)$$

$$\mu_2 = \min(\rho \mu_2, \mu_{\max}) \quad (27)$$

where ρ is the gain coefficient larger than 1. Finally, the iteration is stopped when the maximum iteration number is reached or the minimum reconstructed error $\|\mathbf{L} - \mathbf{H} \mathbf{B} \mathbf{R}\|_F^2 / \|\mathbf{L}\|_F^2$ is smaller than $\delta = 10^{-10}$. The iteration scheme of the optimization algorithm is summarized in Algorithm 1.

Algorithm 1: ADMM for Solving the Fusion Method

Input: PAN image \mathbf{P} , LR MS image \mathbf{L} , downsampling matrix \mathbf{R} , blur matrix \mathbf{B} , spectral response matrix \mathbf{S} , parameters L, h, s, α, β , and λ ;

Initialization: $\mathbf{H}^0 = \mathbf{X}^0 = \mathbf{0}$, $\mathbf{Y}_1 = \mathbf{0}$, $\mathbf{Y}_2 = \mathbf{0}$, $\mu_1^0 = \mu_2^0 = 10^{-3}$, $\rho = 1.1$, $\mu_{\max} = 10^{15}$, $\delta = 10^{-10}$, $maxIter = 30$, $t = 0$;

Training dictionary \mathbf{D} from \mathbf{P}^T ;

while $\|\mathbf{L} - \mathbf{H} \mathbf{B} \mathbf{R}\|_F^2 / \|\mathbf{L}\|_F^2 > \delta$ and $t < maxIter$ **do**

1. Update \mathbf{H}^{t+1} via (16);

2. Update \mathbf{X}^{t+1} via (21);

3. **for** $i = 1$ to $4K$ **do**

Update \mathbf{a}_i^{t+1} via (23) by LARS;

end for;

4. Update \mathbf{Y}_1^{t+1} and \mathbf{Y}_2^{t+1} via (24) and (25);

5. Update μ_1^{t+1} and μ_2^{t+1} via (26) and (27);

6. $t \leftarrow t + 1$;

end

Output: HR MS image \mathbf{H} .

V. EXPERIMENTS

In this section, the experiment datasets are introduced in detail. In addition, parameter settings are also given. Then, we analyze the fusion results of the proposed method and all compared methods on different datasets. In addition, some important parameters are also further investigated for a comprehensive understanding.

A. Experiment Setup

In the experiment section, different datasets from QuickBird and GeoEye-1 satellites are employed and fused. Because there is no HR MS image, also termed as the reference image, for a direct comparison, reduced-resolution datasets are produced by downsampling and blurring from original source images in order to verify the reconstruction performance. For example, the $64 \times 64 \times 4$ LR MS and 256×256 PAN images in Fig. 2(a) and (b) are generated from $256 \times 256 \times 4$ HR MS and 1024×1024 HR PAN images by downsampling with ratio 4 and blurring, in which the blurring kernels are constructed by the settings in [56]. The images in Fig. 2(c) and (d) are produced in the same way. Then, the $256 \times 256 \times 4$ fused images of reduced-resolution datasets are directly compared with the original HR MS images. There are four indexes for quality assessment of reduced-resolution datasets, including Q4 [57], spectral angle mapper (SAM) [58], universal image quality index (UIQI) [59], and *Erreur Relative Globale Adimensionnelle de Synthèse* (ERGAS) [60]. Q4 and UIQI vary from 0 to 1 and the fusion results will be better for larger values. For SAM and ERGAS, smaller values mean better fusion results and the best value is 0. Besides, the fusion experiments are also conducted on the full-resolution datasets in Fig. 2(e)–(h), where the sizes of LR MS and PAN images are $64 \times 64 \times 4$ and 256×256 , respectively. For full-resolution datasets, there are no reference images for comparisons. Thus, some no-reference

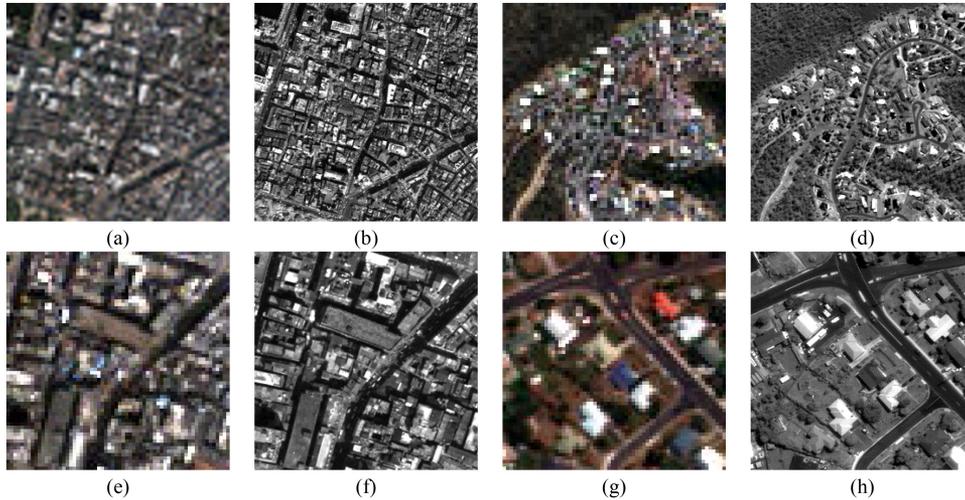


Fig. 2. Source images to be fused. (a) and (b) Reduced-resolution LR MS and PAN images from QuickBird. (c) and (d) Reduced-resolution LR MS and PAN images from GeoEye-1. (e) and (f) Full-resolution LR MS and PAN images from QuickBird. (g) and (h) Full-resolution LR MS and PAN images from GeoEye-1.

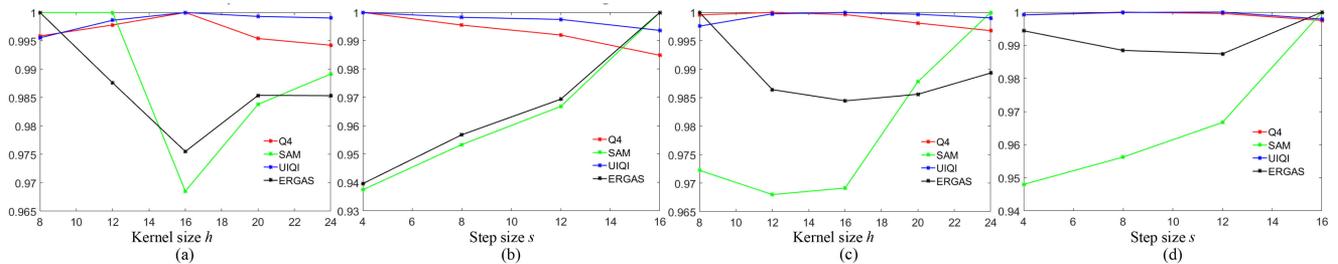


Fig. 3. Influences of the kernel size and step size on the fusion results of the reduced-resolution GeoEye-1 and QuickBird datasets. (a) and (b) Kernel size and step size on GeoEye-1 satellite. (c) and (d) Kernel size and step size on QuickBird satellite.

image quality indexes are adopted for assessment, such as D_S , D_λ , and QNR [61]. The best values of D_S and D_λ are 0. For QNR, larger values mean better fused results. Besides, several methods are employed for comparison, including generalized intensity-hue-saturation (GIHS) [8], Gram–Schmidt (GS) [10], proportional additive wavelet LHS (AWLP) [13], Indusion [18], simultaneous image registration and fusion (SIRF) [27], low-rank pansharpening (LRP) [46], pansharpening by convolution neural networks (PNN) [32], deep residual pansharpening neural network (DRPNN) [37], and PanNet [35].

B. Implementation Settings

This section gives the parameter settings in detail. In the proposed method, several parameters, such as α , β , and λ , have important influences on the fused results. For QuickBird dataset, α , β , and λ are set as 2^7 , 2^3 , and 0.1, respectively. The parameters α , β , and λ are set as 2^{10} , 2^5 , and 0.01 for GeoEye-1 dataset. Taking the reconstruction performance and complexity of the proposed method into consideration, $h_1 = h_2 = 16$ and $s_1 = s_2 = 8$ are adopted for the formulation of the convolution matrix. The number of atoms L in the dictionary is set as 100. For a comprehensive understanding of the proposed method,

the influences of these parameters are analyzed in the following sections.

C. Investigation on Kernel Size and Step Size in SCM

This section presents the influences of the kernel and step sizes on the fusion results. Here, we consider the reduced-resolution datasets from GeoEye-1 and QuickBird satellites. Fig. 3 demonstrates the variations of all indexes in terms of different kernel and step sizes. In Fig. 3(a), the kernel size increases from 8 to 24 with step size 4 and we can find the values vary with the increasing of kernel size. Small kernel size will result in much more complexity compared with other sizes. For large kernel size, spatial effects will be introduced. The best values of all indexes are produced when the kernel size equals 16. In Fig. 3(b), the step size varies from 4 to 16 with step 4. It can be observed that small step size can produce better values for all indexes. When the step size is 4 for PAN image, LR MS image is transformed into CM, because the step size for LR MS image is 1. However, the running time for step size 4 is more than about 43 times that of step size 8. Thus, the step size is set as 8 in the proposed method by considering the computational complexity for GeoEye-1 dataset. Besides, we can see similar

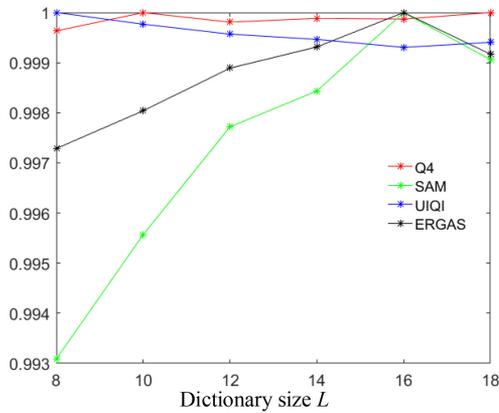


Fig. 4. Influences of dictionary size on the fusion result of reduced-resolution GeoEye-1 dataset.

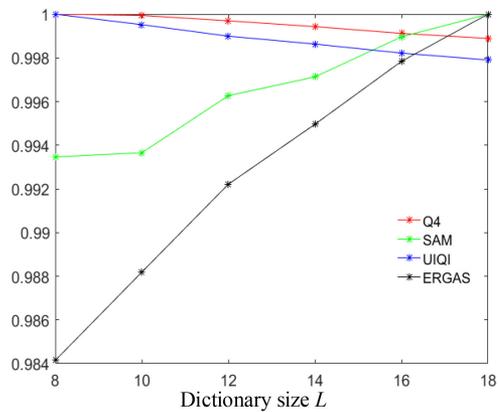


Fig. 5. Influences of dictionary size on the fusion result of reduced-resolution QuickBird dataset.

trends in Fig. 3(c) and (d). The kernel size and step size for QuickBird dataset are also set as 16 and 8, respectively.

D. Investigation on Dictionary Size

The influences of dictionary size on fusion results are analyzed in this section. The values of all indexes are normalized for consistent presentation in Figs. 4 and 5. The dictionary size L varies from 8^2 to 18^2 with step size 2 on the base number. In Figs. 4 and 5, we can see that dictionary size has less effect on the fusion results, although the values of all indexes generally become better with the increase of dictionary size. However, the complexity of the proposed method will increase with increasing dictionary size. Therefore, the number of atoms in the dictionary is set as 100 through a comprehensive evaluation.

E. Investigation on Regularization Parameters

In this section, we analyze the influences of regularization parameters on the fusion results. The experiment is conducted on the image pair on the reduced-resolution data of GeoEye-1 satellite. Fig. 6 displays the results of the image pair, where the values on axes are the logs base 2 of α and β . From Fig. 6,

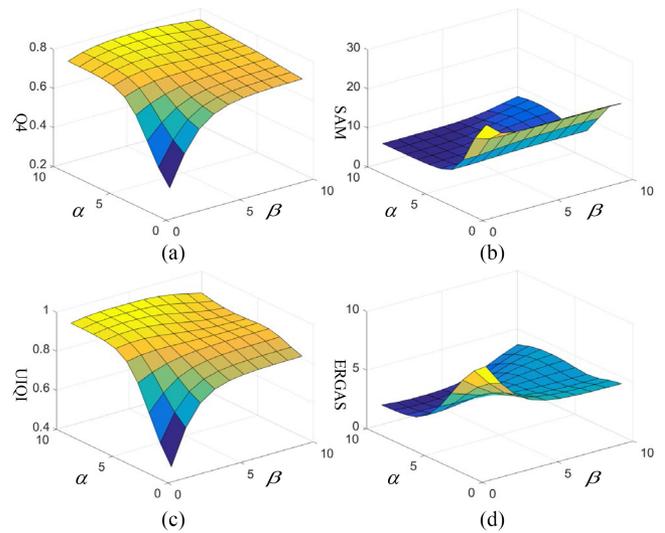


Fig. 6. Influences of α and β on fusion result of reduced-resolution GeoEye-1 dataset.

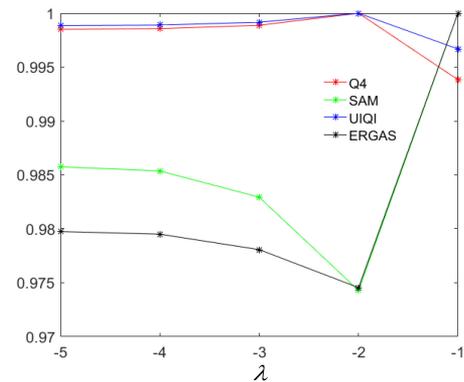


Fig. 7. Influences of λ on fusion result of reduced-resolution GeoEye-1 dataset.

one can observe that the index values vary with different α and β . Q4 and UIQI decrease with the increasing of α and β . For SAM, α has a more obvious impact on index values than β . ERGAS has a similar trend with SAM. Besides, we can find that the best values for all indexes cannot be achieved on the same settings. Thus, α and β are set as 2^{10} and 2^5 by considering the overall performance on GeoEye-1 dataset. Besides, we also investigate the influences of λ on the fusion results. For a uniform presentation, the values of all indexes are normalized and illustrated in Fig. 7, in which the values on axes are the logs base 10 of λ . From Fig. 7, it can be seen that the best values are obtained when λ equals 0.01 for all indexes. So, the images from GeoEye-1 satellite are fused with the settings. By similar experiments, the settings for QuickBird satellite dataset are found.

F. Experiments on Reduced-Resolution Datasets

The experimental results on the reduced-resolution datasets are displayed and analyzed. Fig. 8 shows the fused images of

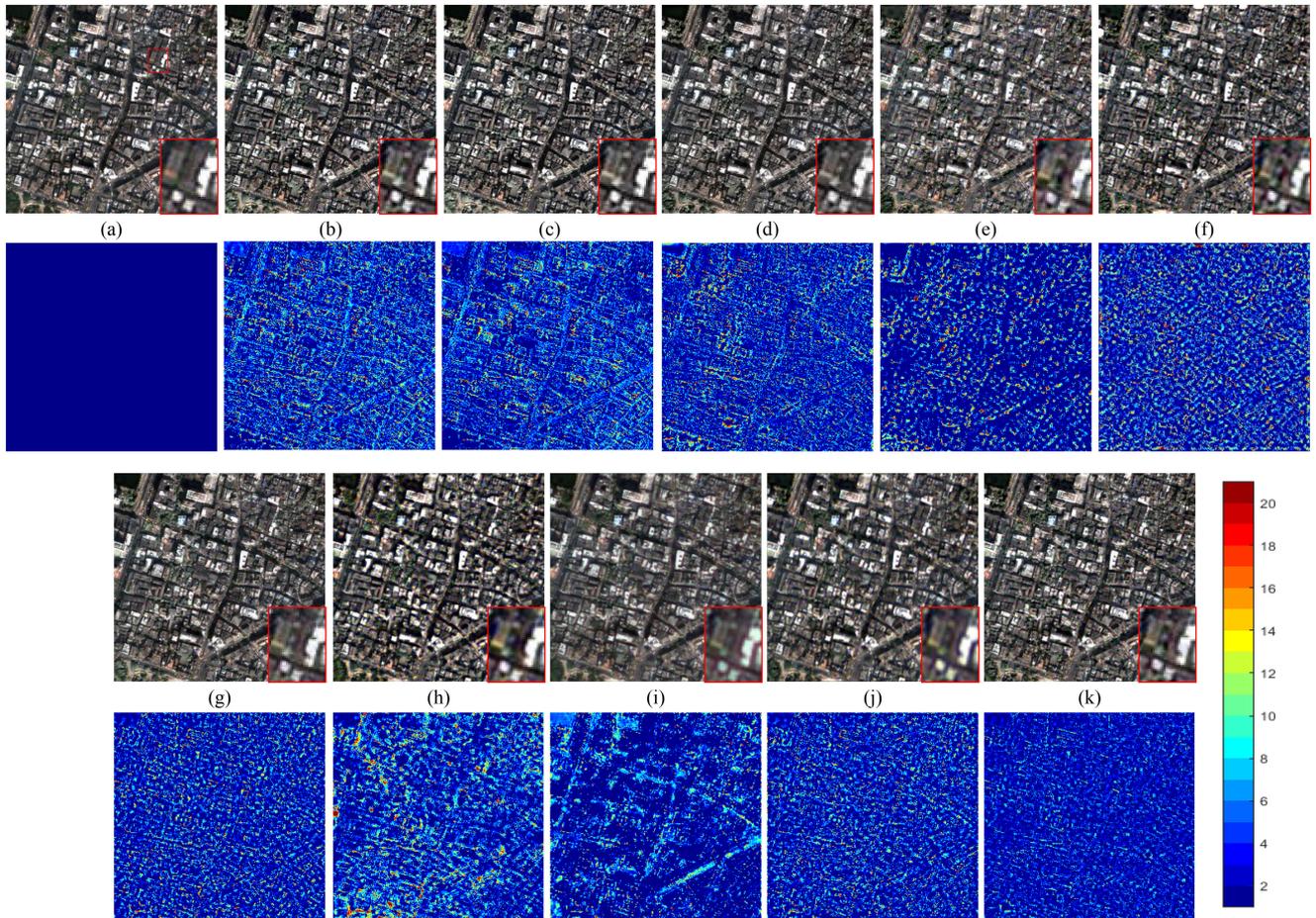


Fig. 8. Fusion results of images from QuickBird. (a) Reference image. (b) GIHS [8]. (c) GS [10]. (d) AWLP [13]. (e) Indusion [18]. (f) SIRF [27]. (g) LRP [46]. (h) PNN [32]. (i) DRPNN [37]. (j) PanNet [35]. (k) Proposed method.

all methods on QuickBird dataset. Besides, we display the error maps between fused images and the reference image for more intuitive analysis. The reference image is placed in Fig. 8(a) and all fused images are compared with it. From the images in Fig. 8, it can be seen that spectral distortions appear in Fig. 8(b) and (c), which may be caused by the spectral range differences in PAN and MS images for component substitution-based methods. But clear spatial information is provided by Fig. 8(b) and (c). The spectral distortions can also be found in Fig. 8(d), especially in the vegetation areas. In Fig. 8(e), we can see some obvious spatial artifacts, which result from the decimation in Indusion [18]. Similar spatial effects also can be observed in Fig. 8(h) and spatial blurring is introduced. The spectral features in Fig. 8(i) are preserved better than Fig. 8(g), but some differences still exist. For the proposed method, the spatial and spectral information in Fig. 8(k) is more consistent with that of the reference in Fig. 8(a). Besides, a local area is framed by a red rectangle and enlarged for a closer inspection. The enlarged areas are put on the bottom right corner of the fused images. From these areas, one can see that some subtle spectral information is lost in Fig. 8(b)–(e). In the amplified area in Fig. 8(g), the spectral features are promoted excessively. The spectral information is preserved well in the local area of Fig. 8(k). Besides, we can

observe that the fusion performance of the proposed method is better than that of the other methods from the error maps of the fused images.

Table I lists the values of all evaluation indexes and the best value for each index is labeled in bold. In Table I, we can see that best SAM, UIQI, and ERGAS are from the result of the proposed method, but AWLP [13] provides the best Q4. The proposed method can achieve a better performance generally.

Fig. 9 shows the fusion results of all methods on GeoEye-1 dataset. In Fig. 9, it can be observed that spectral differences in Fig. 9(b) and (c) are considerable compared with the reference image in Fig. 9(a). The color tone of both images is different, although GIHS [8] and GS [10] belong to the same kind of methods. In the vegetation areas, there are some differences in Fig. 9(d) and (e), but the color for the building areas is maintained. Inversely, the spectral features in Fig. 9(f) are distorted and the color of some vegetation areas becomes gray. The spatial details are enhanced well in Fig. 9(g), but some blurring effects should be noticed in Fig. 9(h), especially in some edges of buildings. Fig. 9(k) behaves better in spatial and spectral preservation. In addition, an interesting region is selected and enlarged for comparison. In the amplified area of Fig. 9(e), the spatial artifacts can be observed obviously. The

TABLE I
NUMERICAL EVALUATION OF FUSED RESULTS ON REDUCED-RESOLUTION QUICKBIRD DATASET

Evaluation index	GIHS [8]	GS [10]	AWLP [13]	Indusion [18]	SIRF [27]	LRP [46]	PNN [32]	DRPNN [37]	PanNet [35]	Proposed method
Q4	0.8737	0.8682	0.8840	0.8200	0.8523	0.8817	0.8558	0.8553	0.8826	0.8738
SAM	10.4406	9.9440	10.1127	10.1260	11.1616	9.0019	14.9718	9.0824	11.4630	8.6511
UIQI	0.9067	0.8953	0.9180	0.8822	0.9035	0.9266	0.9225	0.9096	0.9176	0.9347
ERGAS	3.4446	3.6700	3.2040	3.8135	3.4007	2.8905	3.8884	2.9797	3.0227	2.7899

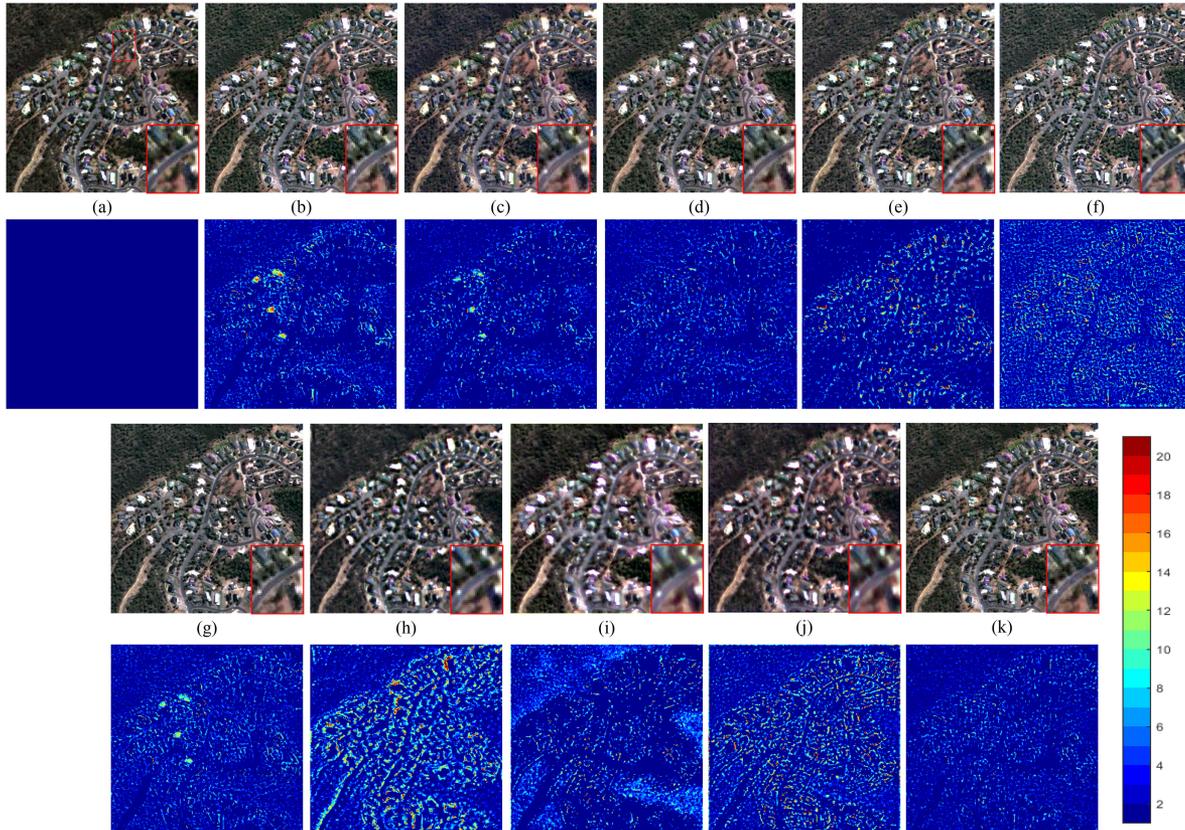


Fig. 9. Fusion results of images from GeoEye-1. (a) Reference image. (b) GIHS [8]. (c) GS [10]. (d) AWLP [13]. (e) Indusion [18]. (f) SIRF [27]. (g) LRP [46]. (h) PNN [32]. (i) DRPNN [37]. (j) PanNet [35]. (k) Proposed method.

TABLE II
NUMERICAL EVALUATION OF FUSED RESULTS ON REDUCED-RESOLUTION GEOEYE-1 DATASET

Evaluation index	GIHS [8]	GS [10]	AWLP [13]	Indusion [18]	SIRF [27]	LRP [46]	PNN [32]	DRPNN [37]	PanNet [35]	Proposed method
Q4	0.7224	0.7103	0.7648	0.7599	0.7990	0.7969	0.7645	0.7808	0.7655	0.7706
SAM	8.9410	5.8790	5.7086	5.0248	5.0642	5.2243	6.3737	6.6887	5.9017	4.9843
UIQI	0.8479	0.8697	0.9223	0.9189	0.9329	0.9230	0.9242	0.9173	0.9297	0.9409
ERGAS	4.4562	3.9941	1.6112	1.7288	1.6371	2.1585	1.8983	2.0793	2.1859	1.5355

spectral information is lost in the local area of Fig. 9(f). Besides, the enlarged area of Fig. 9(k) is more consistent with that of the reference image visually.

The numerical values of all indexes are provided in Table II and the best values are highlighted. In Table II, we can see that the best values of SAM, UIQI, and ERGAS are provided by the proposed method. Although SIRF [27] provides the best Q4, the proposed method also behaves better than the component

substation-based methods and PNN [32]. From the error maps, it can be observed that the errors of GIHS [8], GS [10], and LRP [46] are obvious in some vegetation regions

G. Experiments on Full-Resolution Datasets

This section presents the fusion results of full-resolution datasets from QuickBird and GeoEye-1 satellites. Fig. 10

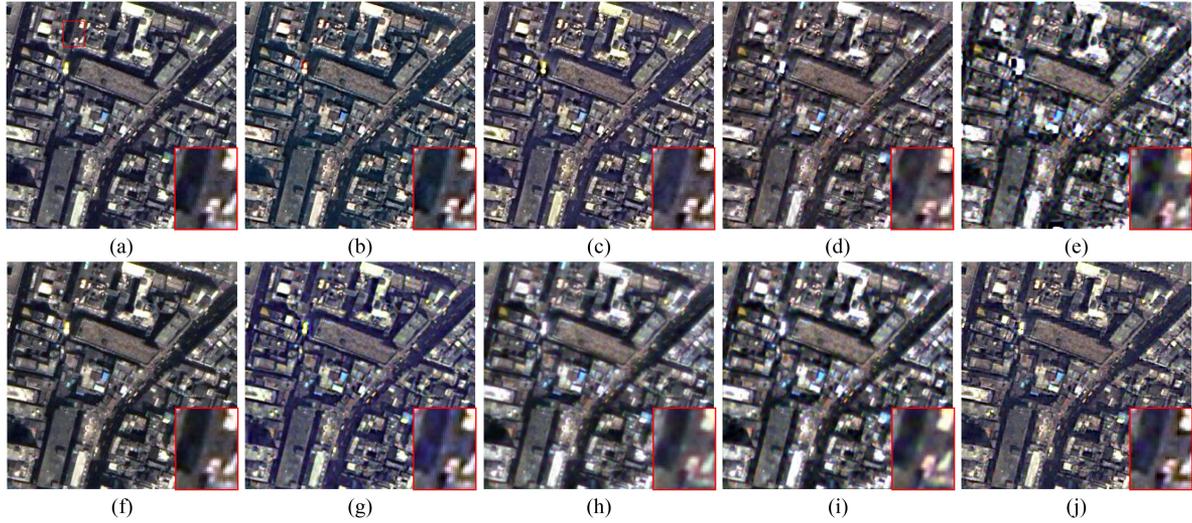


Fig. 10. Fusion results of images from QuickBird. (a) GIHS [8]. (b) GS [10]. (c) AWLP [13]. (d) Indusion [18]. (e) SIRF [27]. (f) LRP [46]. (g) PNN [32]. (h) DRPNN [37]. (i) PanNet [35]. (j) Proposed method.

TABLE III
NUMERICAL EVALUATION OF FUSED RESULTS ON FULL-RESOLUTION QUICKBIRD DATASET

Evaluation index	GIHS [8]	GS [10]	AWLP [13]	Indusion [18]	SIRF [27]	LRP [46]	PNN [32]	DRPNN [37]	PanNet [35]	Proposed method
D_S	0.1416	0.1254	0.1103	0.1939	0.2297	0.1257	0.0398	0.0562	0.0476	0.0324
D_λ	0.1384	0.1256	0.1433	0.0089	0.0248	0.1190	0.0822	0.0353	0.0473	0.0332
QNR	0.7396	0.7646	0.7623	0.7989	0.7645	0.7702	0.8813	0.9105	0.9073	0.9355

demonstrates the fused images of all methods on QuickBird dataset. In Fig. 10, we can see that the spatial details are enhanced well for all images but there still exist some differences in hue for different fused images. The spectral features in Fig. 10(a) are more similar to those in Fig. 10(c), although Fig. 10(a) and (b) is derived from the same framework. It can be observed that the spatial information of Fig. 10(e) has obvious distortions compared with the results of other methods. Besides, some spatial effects also arise. The same performance can be found in Fig. 10(d) and the color of some buildings seems unnatural. For example, the color in the local area slightly becomes red. In addition, some spectral artifacts also appear in Fig. 10(g) and the spectral features in the amplified area of Fig. 10(g) are oversaturated. Compared with other fused images, the proposed method can produce more natural color for buildings when improving the spatial details. The evaluation values of all indexes are reported in Table III. From Table III, one can see that the best value of D_S is achieved by the proposed method, which shows the effectiveness in spatial information. Although the best value of D_λ is from Indusion [18], the proposed method behaves better in the overall index QNR.

Fig. 11 exhibits the fusion results of all methods on GeoEye-1 dataset. We can see some color differences on road between Fig. 11(a) and (b), which may be caused by different spectral response settings. For the result in Fig. 11(c), the color of the roof in the upper right corner is different from that in Fig. 10(b). Severe spatial effects cannot be ignored in Fig. 11(d). However,

spatial blurring effects can be observed in Fig. 11(g). The result in Fig. 11(j) has a better performance in spatial and spectral information. Moreover, the detailed analysis in the amplified areas also has a similar performance. For example, some spatial distortions, such as checkerboard artifacts, are noticeable. The edges and textures are blurred in the enlarged area of Fig. 11(g). From the local area in Fig. 11(j), we can see clear spatial details. Table IV illustrates the numerical values of all fused images and the best values are labeled in bold. In Table IV, the best values of D_S and QNR are from the proposed method, which is consistent with the visual analysis in Fig. 11.

H. Complexity Analysis and Comparison

The complexity of the proposed method is analyzed as follows. In the optimization algorithm, the updates of \mathbf{H} and \mathbf{X} dominate the computational complexity. For \mathbf{H} , the Sylvester equation is solved, whose complexity is $o((4Kh_1h_2)^3)$ due to the computation of matrix inversion. The update of \mathbf{X} involves SVD operation and its complexity is $o(2(l_1 - h_1 + s_1)(l_2 - h_2 + s_2)h_1^2h_2^2/s_1s_2)$. Besides, for the estimation of \mathbf{a}_i , its complexity is similar to that of the least squares method in the worst case scenario, but it most often has a very good performance in practical ones.

In Table V, the average running time of all methods is illustrated, in which the time is recorded in seconds. The first six methods and the proposed method are tested on the same computer with Intel Core i7-6700 processor, 3.4GHz, Intel

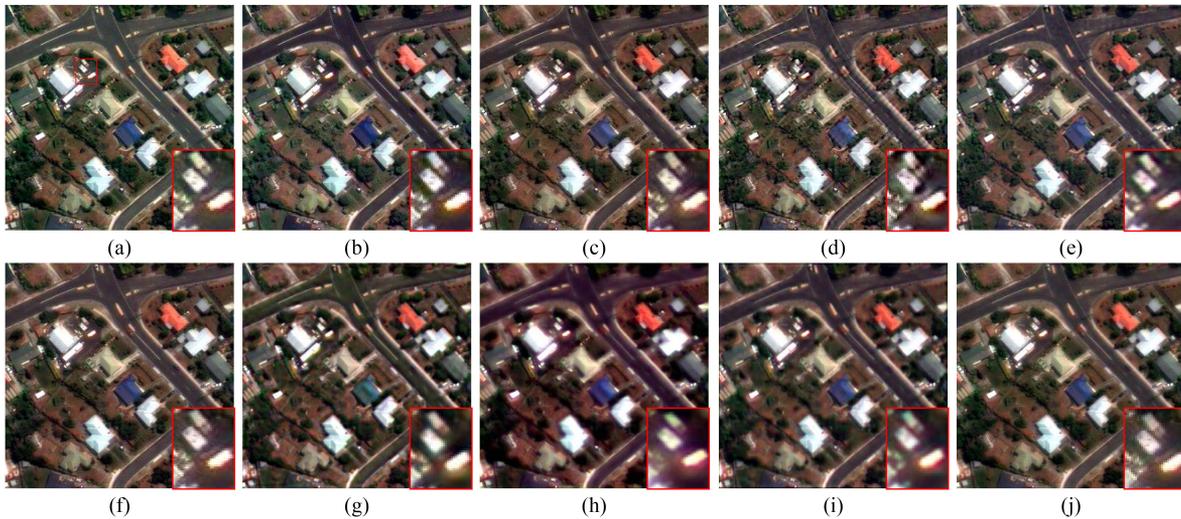


Fig. 11. Fusion results of images from GeoEye-1. (a) GIHS [8]. (b) GS [10]. (c) AWLP [13]. (d) Indusion [18]. (e) SIRF [27]. (f) LRP [46]. (g) PNN [32]. (h) DRPNN [37]. (i) PanNet [35]. (j) Proposed method.

TABLE IV
NUMERICAL EVALUATION OF FUSED RESULTS ON FULL-RESOLUTION GEOEYE-1 DATASET

Evaluation index	GIHS [8]	GS [10]	AWLP [13]	Indusion [18]	SIRF [27]	LRP [46]	PNN [32]	DRPNN [37]	PanNet [35]	Proposed method
D_s	0.1129	0.1244	0.0930	0.0492	0.0677	0.0735	0.0538	0.0884	0.0878	0.0273
D_λ	0.1079	0.1261	0.1347	0.1332	0.0800	0.1013	0.0800	0.0723	0.0894	0.0789
QNR	0.7913	0.7652	0.7849	0.8242	0.8577	0.8326	0.8705	0.8411	0.8306	0.8959

TABLE V
TIME COMPARISON OF ALL METHODS

Method	Time
GIHS [8]	0.77
GS [10]	0.81
AWLP [13]	0.53
Indusion [18]	0.79
SIRF [27]	48.60
LRP [46]	2.87
PNN [32]	4.85
DRPNN [37]	3.05
PanNet [35]	10.98
Proposed method	434.10

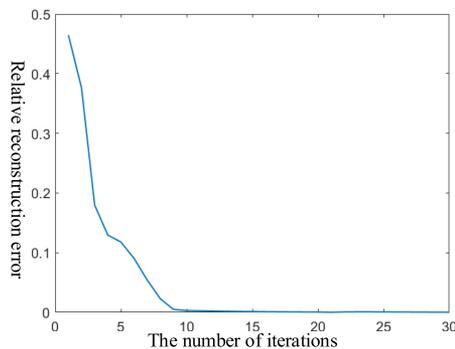


Fig. 12. Convergence curve of the optimization algorithm based on ADMM.

UHD Graphics 630 and 16G memory by MATLAB R2017a. From Table V, we can see that component substitution- and MRA-based methods spend little time for fusion when compared with other kinds of methods. For PNN [32] and DRPNN [37], their MATLAB codes are provided and implemented on the same computer mentioned above for test, where MatConvNet toolbox is compiled for running. PanNet [35] is trained and tested by PyTorch on an NVIDIA 2080Ti GPU with Intel Core i7-9700 processor, 3.0GHz and 128 G memory. The testing time for DNN-based pansharpening methods is fast but their training may take a long time. Because iterative optimization is used in the proposed method, the running time is considerable.

I. Convergence Analysis

In this section, we analyze the convergence of the proposed optimization algorithm based on ADMM and plot the convergence curve in Fig. 12. From Fig. 12, one can observe that the algorithm derived from ADMM can converge rapidly, which ensures the performance of the proposed method.

VI. CONCLUSION

In this article, we proposed a novel fusion method for LR MS and PAN images based on low-rank and sparse SCM. SCM is formulated to reduce the computation complexity of CM and capture the local structures in MS images. Then, the degraded

relationships of LR MS and PAN images from the HR MS images are established, which are then incorporated with the low-rank prior of SCM. Next, the fused HR MS image is sparsely represented over a dictionary containing local structures, which is learned from the corresponding PAN image. Finally, the fusion model is solved through ADMM. Compared with GIHS [8], GS [10], AWLP [13], Indusion [18], SIRF [27], LRP [46], PNN [32], DRPNN [37], and PanNet [35], the proposed method can produce better fusion results in visual analysis and achieve a better performance in numerical evaluation. Owing to the introduction of low-rank and sparse properties of SCM, the proposed method can efficiently preserve the spatial and spectral structures in the fused images.

REFERENCES

- [1] X. Qin, J. Shen, X. Mao, X. Li, and Y. Jia, "Robust match fusion using optimization," *IEEE Trans. Cybern.*, vol. 45, no. 8, pp. 1549–1560, Aug. 2015.
- [2] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Hyperspectral images super-resolution via learning high-order coupled tensor ring representation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 11, pp. 4747–4760, Nov. 2020.
- [3] L. Zhang, H. Shen, W. Gong, and H. Zhang, "Adjustable model-based fusion method for multispectral and panchromatic images," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 6, pp. 1693–1704, Dec. 2012.
- [4] J. Han, H. Chen, N. Liu, C. Yan, and X. Li, "CNNs-based RGB-D saliency detection via cross-view transfer and multiview fusion," *IEEE Trans. Cybern.*, vol. 48, no. 11, pp. 3171–3183, Nov. 2018.
- [5] G. Vivone *et al.*, "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, May 2015.
- [6] P. Liu and L. Xiao, "Multicomponent driven consistency priors for simultaneous decomposition and pansharpening," *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, vol. 12, no. 11, pp. 4589–4605, Nov. 2019.
- [7] X. Ma, X. Zhang, X. Tang, H. Zhou, and L. Jiao, "Hyperspectral anomaly detection based on low-rank representation with data-driven projection and dictionary construction," *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, vol. 13, pp. 2226–2239, 2020.
- [8] T. M. Tu, P. S. Huang, C. L. Hung, and C. P. Chang, "A fast intensity-hue-saturation fusion technique with spectral adjustment for IKONOS imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 4, pp. 309–312, Oct. 2004.
- [9] P. S. Chavez, Jr., S. C. Sides, and J. A. Anderson, "Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic," *Photogramm. Eng. Remote Sens.*, vol. 57, no. 3, pp. 295–303, Mar. 1991.
- [10] C. A. Laben and B. V. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," U.S. Patent 6011875, Jan. 4, 2000.
- [11] B. Aiuzzi, S. Baronti, and M. Selva, "Improving component substitution pansharpening through multivariate regression of MS+Pan data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3230–3239, Oct. 2007.
- [12] 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. 79, no. 1, pp. 1–22, Sep. 2011.
- [13] X. Otazu, M. González-Audicana, O. Fors, and J. Núñez, "Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 10, pp. 2376–2385, Oct. 2005.
- [14] 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.
- [15] M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2096–2106, Dec. 2005.
- [16] M. Choi, R. Y. Kim, M. R. Nam, and H. O. Kim, "Fusion of multispectral and panchromatic satellite images using the curvelet transform," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 2, pp. 136–140, Apr. 2005.
- [17] J. L. Starck, E. J. Candès, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Process.*, vol. 11, no. 6, pp. 670–684, Jun. 2002.
- [18] M. M. Khan, J. Chanussot, L. Condat, and A. Montavert, "Indusion: Fusion of multispectral and panchromatic images using the induction scaling technique," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 1, pp. 98–102, Jan. 2008.
- [19] B. Aiuzzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "MTF tailored multiscale fusion of high-resolution MS and pan imagery," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 5, pp. 591–596, May 2006.
- [20] Y. Xing, M. Wang, S. Yang, and K. Zhang, "Pansharpening with multiscale geometric support tensor machine," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2503–2517, May 2018.
- [21] M. Guo, H. Zhang, J. Li, L. Zhang, and H. Shen, "An online coupled dictionary learning approach for remote sensing image fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1284–1294, Apr. 2014.
- [22] K. Zhang, M. Wang, S. Yang, and L. Jiao, "Convolution structure sparse coding for fusion of panchromatic and multispectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1117–1130, Feb. 2019.
- [23] W. Wang, L. Jiao, and S. Yang, "Fusion of multispectral and panchromatic images via sparse representation and local autoregressive model," *Inf. Fusion*, vol. 20, pp. 73–87, Nov. 2014.
- [24] H. Yin, "PAN-guided cross-resolution projection for local adaptive sparse representation based pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4938–4950, Jul. 2019.
- [25] K. Zhang, M. Wang, S. Yang, Y. Xing, and R. Qu, "Fusion of panchromatic and multispectral images via coupled sparse non-negative matrix factorization," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 12, pp. 5740–5747, Dec. 2016.
- [26] M. Lotfi and H. Ghassemian, "A new variational model in texture space for pansharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 8, pp. 1269–1273, Aug. 2018.
- [27] 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.
- [28] L. Deng, M. Feng, and X. C. Tai, "The fusion of panchromatic and multispectral remote sensing images via tensor-based sparse modeling and hyper-Laplacian prior," *Inf. Fusion*, vol. 52, pp. 76–89, Dec. 2019.
- [29] W. Huang, L. Xiao, Z. Wei, H. Liu, and S. Tang, "A new pan-sharpening method with deep neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 1037–1041, May 2015.
- [30] Z. Shao and J. Cai, "Remote sensing image fusion with deep convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1656–1669, May 2018.
- [31] Y. Zhang, C. Liu, M. Sun, and Y. Ou, "Pan-sharpening using an efficient bidirectional pyramid network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5549–5563, Aug. 2019.
- [32] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, 2016, Art. no. 594.
- [33] X. Liu, Q. Liu, and Y. Wang, "Remote sensing image fusion based on two-stream fusion network," *Inf. Fusion*, vol. 55, pp. 1–15, Mar. 2020.
- [34] J. Ma, W. Yu, C. Chen, P. Liang, X. Guo, and J. Jiang, "Pan-GAN: An unsupervised pan-sharpening method for remote sensing image fusion," *Inf. Fusion*, vol. 62, pp. 110–120, Oct. 2020.
- [35] J. Yang, X. Fu, Y. Hu, Y. Huang, X. Ding, and J. Paisley, "PanNet: A deep network architecture for pan-sharpening," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 5449–5457.
- [36] Y. Fu, W. Wang, Y. Huang, X. Ding, and J. Paisley, "Deep multi-scale detail networks for multi-band spectral image sharpening," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: [10.1109/TNNLS.2020.2996498](https://doi.org/10.1109/TNNLS.2020.2996498).
- [37] Y. Wei, Q. Yuan, H. Shen, and L. Zhang, "Boosting the accuracy of multispectral image pan-sharpening by learning a deep residual network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1795–1799, Oct. 2017.
- [38] L. He *et al.*, "Pan-sharpening via detail injection based convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 4, pp. 1188–1204, Apr. 2019.
- [39] L. Deng, G. Vivone, C. Jin, and J. Chanussot, "Detail injection-based deep convolutional neural networks for pan-sharpening," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: [10.1109/TGRS.2020.3031366](https://doi.org/10.1109/TGRS.2020.3031366).
- [40] H. Guo, P. Zhuang, and Y. Guo, "Bayesian pan-sharpening with multi-order gradient-based deep network constraints," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 950–962, 2020.

- [41] S. Mao, L. Xiong, L. Jiao, T. Feng, and S. Yeung, "A novel Riemannian metric based on Riemannian structure and scaling information for fixed low-rank matrix completion," *IEEE Trans. Cybern.*, vol. 47, no. 5, pp. 1299–1312, May 2017.
- [42] Z. Li, F. Nie, X. Chang, L. Nie, H. Zhang, and Y. Yang, "Rank-constrained spectral clustering with flexible embedding," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 12, pp. 6073–6082, Dec. 2018.
- [43] X. Dong *et al.*, "Weighted locality collaborative representation based on sparse subspace," *J. Vis. Commun. Image Representation*, vol. 58, pp. 187–194, Jan. 2019.
- [44] X. He, L. Condat, J. M. Bioucas-Dias, J. Chanussot, and J. Xia, "A new pansharpening method based on spatial and spectral sparsity priors," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 4160–4174, Sep. 2014.
- [45] 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.
- [46] S. Yang, K. Zhang, and M. Wang, "Learning low-rank decomposition for pan-sharpening with spatial-spectral offsets," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 8, pp. 3647–3657, Aug. 2018.
- [47] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.
- [48] G. Liu, S. Chang, and Y. Ma, "Blind image deblurring using spectral properties of convolution operators," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5047–5056, Dec. 2014.
- [49] G. Liu and W. Zhang, "Recovery of future data via convolution nuclear norm minimization," 2019, *arXiv:1909.03889*.
- [50] J. Cai, E. Candes, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [51] Y. Peng, J. Suo, Q. Dai, and W. Xu, "Reweighted low-rank matrix recovery and its application in image restoration," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2418–2430, Dec. 2014.
- [52] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Nonlocal patch tensor sparse representation for hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 28, no. 6, pp. 3034–3047, Jun. 2019.
- [53] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, Montreal, QC, Canada, 2009, pp. 689–696.
- [54] 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.
- [55] J. Mairal, F. Bach, and J. Ponce, "Sparse modeling for image and vision processing," *Found. Trends Comp. Graph. Vis.*, vol. 8, pp. 85–283, 2014.
- [56] M. M. Khan, L. Alparone, and J. Chanussot, "Pansharpening quality assessment using the modulation transfer functions of instruments," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3880–3891, Nov. 2009.
- [57] L. Alparone, S. Baronti, A. Garzelli, and F. Nencini, "A global quality measurement of pan-sharpened multispectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 4, pp. 313–317, Oct. 2004.
- [58] R. H. Yuhas, A. F. H. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm," in *Proc. Summaries 3rd Annu. JPL Airborne Geosci. Workshop*, 1992, pp. 147–149.
- [59] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.
- [60] 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.
- [61] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva, "Multispectral and panchromatic data fusion assessment without reference," *Photogramm. Eng. Remote Sens.*, vol. 74, no. 2, pp. 193–200, Feb. 2008.



Feng Zhang received the B.S. degree in electronic information engineering from Shandong Normal University, Ji'nan, China, in 2012 and the M.Sc. degree in electronic information engineering from Xidian University, Xi'an, China, in 2016. She is currently working toward the Ph.D. degree in computer science and technology with the School of Information Science and Engineering, Shandong Normal University.

Her main current research interests include deep learning and image processing.



Haoran Zhang is currently working toward the bachelor's degree in computer science and technology with the School of Information Science and Engineering, Shandong Normal University, Ji'nan, China, under the supervision of K. Zhang.

His current research interest includes remote sensing image fusion.



Kai Zhang was born in Shanxi, China, in 1992. He received the B.S. degree in electrical engineering and automation from the North University of China, Taiyuan, China, in 2013 and the Ph.D. degree in circuit and system from Xidian University, Xi'an, China, in 2018.

He is currently a Lecturer with the School of Information Science and Engineering, Shandong Normal University, Ji'nan, China. His research interests include multisource remote sensing image fusion, matrix factorization, and deep learning.



Yinghui Xing received the B.S. and Ph.D. degrees in circuit and system from the School of Artificial Intelligence, Xidian University, Xi'an, China, in 2014 and 2020, respectively.

She is an Associate Professor with the School of Computer Science, Northwestern Polytechnical University, Xi'an, China. Her research interests include remote sensing image processing, image fusion, and image super-resolution.



Jiande Sun received the Ph.D. degree in communication and information system from Shandong University, Ji'nan, China, in 2000 and 2005, respectively.

From 2008 to 2009, he was a Visiting Researcher with the Institute of Telecommunications System, Technical University of Berlin, Berlin, Germany. From 2010 to 2012, he was a Postdoctoral Researcher with the Institute of Digital Media, Peking University, Beijing, China, and also with the State Key Laboratory of Digital-Media Technology, Hisense Group, Qingdao, China. From 2014 to 2015, he was a DAAD

Visiting Researcher with the Technical University of Berlin and the University of Konstanz, Konstanz, Germany. From 2015 to 2016, he was a Visiting Researcher with the School of Computer Science, Language Technology Institute, Carnegie Mellon University, Pittsburgh, PA, USA. He is currently a Professor with the School of Information Science and Engineering, Shandong Normal University. He has authored or coauthored more than 60 journal and conference papers and is the coauthor of two books. His current research interests include multimedia content analysis, video hashing, gaze tracking, image/video watermarking, and 2-D-to-3-D conversion.



Quanyuan Wu was born in 1959. He received the Ph.D. degree in geodesy and surveying engineering from Shandong University of Science and Technology, Qingdao, China, in 2007.

Since 2003, he has been a Professor with the Department of Geography, College of Geography and Environment, Shandong Normal University, Ji'nan, China. His research interests include the development of remote sensing and geographic information techniques.

Dr. Wu is a member of the Education and Science Popularization Committee of the China GIS Association and the Executive Director of the Shandong Remote Sensing Society. He presided over or participated in three Nation Natural Science Foundations of China.