

# Convolution Structure Sparse Coding for Fusion of Panchromatic and Multispectral Images

Kai Zhang<sup>ID</sup>, Min Wang<sup>ID</sup>, Shuyuan Yang<sup>ID</sup>, Senior Member, IEEE, and Licheng Jiao, Fellow, IEEE

**Abstract**—Recently, sparse coding-based image fusion methods have been developed extensively. Although most of them can produce competitive fusion results, three issues need to be addressed: 1) these methods divide the image into overlapped patches and process them independently, which ignore the consistency of pixels in overlapped patches; 2) the partition strategy results in the loss of spatial structures for the entire image; and 3) the correlation in the bands of multispectral (MS) image is ignored. In this paper, we propose a novel image fusion method based on convolution structure sparse coding (CSSC) to deal with these issues. First, the proposed method combines convolution sparse coding with the degradation relationship of MS and panchromatic (PAN) images to establish a restoration model. Then, CSSC is elaborated to depict the correlation in the MS bands by introducing structural sparsity. Finally, feature maps over the constructed high-spatial-resolution (HR) and low-spatial-resolution (LR) filters are computed by alternative optimization to reconstruct the fused images. Besides, a joint HR/LR filter learning framework is also described in detail to ensure consistency and compatibility of HR/LR filters. Owing to the direct convolution on the entire image, the proposed CSSC fusion method avoids the partition of the image, which can efficiently exploit the global correlation and preserve the spatial structures in the image. The experimental results on QuickBird and Geoeye-1 satellite images show that the proposed method can produce better results by visual and numerical evaluation when compared with several well-known fusion methods.

**Index Terms**—Convolution sparse coding (CSC), image fusion, multispectral (MS) image, panchromatic (PAN) image, structure sparsity.

## I. INTRODUCTION

NOWDAYS, more and more panchromatic (PAN) and multispectral (MS) images can be captured in a bundle

Manuscript received February 20, 2018; revised May 23, 2018 and July 13, 2018; accepted August 5, 2018. Date of publication September 10, 2018; date of current version January 21, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 91438103, Grant 61771376, Grant 61771380, Grant 91438201, Grant U1730109, and Grant U1701267, in part by the Equipment Pre-Research Project of the 13th Five-Years Plan under Grant 6140137050206, Grant 414120101026, Grant 614031201013, Grant 6141A020223, Grant 6141B06160301, and Grant 6141B07090102, in part by the Major Research Plan in Shaanxi Province of China under Grant 2017ZDXM-GY-103 and Grant 017ZDCXL-GY-03-02, and in part by the Foundation of the State Key Laboratory of CEMEE under Grant 2018K0101B. (*Corresponding author: Kai Zhang*.)

K. Zhang, S. Yang, and L. Jiao are with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, International Research Center for Intelligent Perception and Computation, Xidian University, Xi'an 710071, China (e-mail: 2428902470@qq.com; syyang2009@gmail.com; lchjiao@mail.xidian.edu.cn).

M. Wang is with the National Key Laboratory of Radar Signal Processing, Xidian University, Xi'an 710071, China (e-mail: wangmin@xidian.edu.cn).

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

Digital Object Identifier 10.1109/TGRS.2018.2864750

by most optical satellites, such as IKONOS, QuickBird, and Geoeye-1. These images have been widely applied in land cover classification [1], change detection [2], and target detection and recognition [3]. Generally, high-spatial-resolution (HR) PAN image with a single broadband channel contains accurate spatial details, whereas low-spatial-resolution (LR) MS image can provide more spectral characteristics because of the physical tradeoff on spatial and spectral resolutions. Therefore, it is difficult to obtain a complete scene interpretation of the observed area only by HR PAN or LR MS image. To take full advantage of the complementary spatial and spectral information, the fusion of HR PAN and LR MS images is proposed to produce HR MS image, which is particularly named as pansharpening.

Due to the flexible property and increasing demand for HR data, pansharpening technique attracted widespread interest and has been developed a lot over the last two decades. Many methods [4]–[6] are proposed to enhance the spatial resolution of LR MS images, which can be divided into three main categories: component substitution (CS) methods, multiresolution analysis (MRA) methods, and model-based methods. For the first category, the classical methods are Gram–Schmidt (GS) [7], principal component analysis [8], and intensity–hue–saturation [9], [10]. These methods employ specific transform to separate the spatial structure and the spectral information in the interpolated LR MS image. Then, the component containing spatial information is replaced by the HR PAN image. Finally, the HR MS image is generated through the corresponding inverse transform. These methods can be implemented easily and behave well in the enhancement of spatial details. But the mismatch in the spectral range between the PAN and MS images results in significant spectral distortions [5].

Methods based on MRA are derived from the *Amélioration de la Résolution Spatiale par Injection de Structures* concept [11], which assumes that the lost spatial details in LR MS image can be inferred from HR PAN image. Thus, these methods extract the spatial details from the HR PAN image and then inject them into the LR MS image to obtain the HR MS image. In order to find more proper high frequencies, some MRA tools, such as wavelet transform [12], [13], contourlet transform [14], and curvelet transform [15], are utilized in the framework. Besides, a pansharpening method based on support value transform (SVT) [16] is also proposed according to the same scheme. MRA methods can efficiently preserve the spectral information of MS images. However, injection coefficients depend on proper models, which maybe cause some spatial effects [17].

For model-based methods, the relationships in the source images and the desired HR MS image are formulated, where HR PAN and LR MS images are regarded as degraded versions of the HR MS image in spectral and spatial domains, respectively. Thus, the fusion task is reformulated as a restoration problem. Then, reasonable priors, such as total variation [18], Bayesian paradigm [19], and nonnegativity [20], are introduced to regularize the restoration model. Recently, sparsity-induced pansharpening methods gain much attention and have been developed a lot. For example, inspired by compressed sensing [21], Li and Yang [22] solved the restoration model with sparse regularization by basis pursuit algorithm [23] to achieve the fusion of HR PAN and LR MS images. This method can well preserve spectral and spatial details, but its application on the full-scale data is limited because the dictionary is constructed by the patches from desired HR MS images. Subsequently, some different dictionary learning strategies [24], [25] are designed to make the formulation more practical. For instance, Li *et al.* [24] constructed the dictionary for HR MS images from the dictionaries of HR PAN and LR MS images by K-SVD algorithm [29]. Different from the framework in [22], the pansharpening method named as SparseFI [26] explored the sparse coding of the MS images over LR/HR dictionary pair constructed from the PAN image, in which the sparse coefficients of the LR MS image are shared with the HR MS image. The method avoids the dictionary construction from HR MS images, but the spectral differences between HR PAN and LR MS images lead to spectral distortions in the fused images. Then, jointly sparse fusion of images (J-SparseFI) [27] is proposed to model the structure correlations among the bands in MS image by joint sparse prior. Besides, Jiang *et al.* [28] presented a two-step sparse coding method with patch normalization (PN-TSSC) method to find the sparse codes of images and the coding residual.

Although sparse coding-based pansharpening methods are widely studied and obtain state-of-the-art performance, some drawbacks still need to be resolved.

- 1) Most of these methods adopt patch-partition-based strategy to process the source images, and then the patches are independently processed without the consideration of consistency constraint for overlapping pixels. The weighted sum of the overlapped pixels from different adjacent patches will bring the ambiguity of aggregation when reconstructing the fused images. The consistency constraint is considered in SparseFI [26], but it is controlled by a weighting factor, which is difficult to determine.
- 2) The divided patches destroy the entire spatial structure of the image. Besides, the atomic structure of the dictionary will be affected to a certain extent because of the redundancy in the overlapped pixels between adjacent patches. Moreover, the redundant features in the dictionary maybe result in the distortion of spatial details.
- 3) Most of the fusion methods based on sparse coding estimate the sparse coefficients of each band in LR MS image successively and independently. However, there exists a strong correlation and similarity among bands

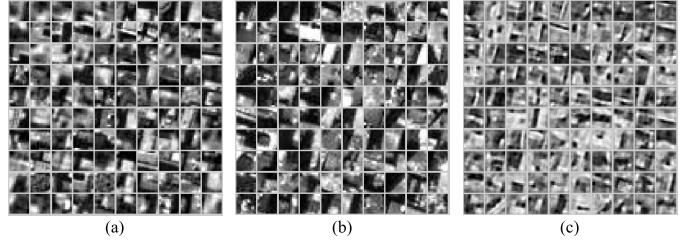


Fig. 1. Visualization of features learned by different methods. (a) Atoms from the dictionary in [24]. (b) Atoms from the dictionary in [26]. (c) Filters learned by CSC.

of MS images. A proper regularizer that captures the structural correlation will improve the performance of the fused images.

To deal with these issues in sparse coding, convolution sparse coding (CSC), as an alternative representation framework, is introduced in [30]–[32] to model an entire image and has important applications in a wide range of computer vision [33]–[35]. CSC can learn the filters from the whole image without patch partition to avoid the redundancy in overlapping pixels, which can well preserve the spatial structure consistency. In order to analyze the learned features, we randomly select 100 atoms of the HR dictionaries trained in [24] and [26] and show them in Fig. 1(a) and (b), respectively. In [24], the dictionary of HR MS image is jointly trained from the LR MS and HR PAN images through K-SVD algorithm. The adopted HR dictionary in [26] is built directly by the HR PAN image. Besides, 100 HR filters are displayed in Fig. 1(c), which are learned by CSC from 50 HR PAN images with a size of  $256 \times 256$  collected by QuickBird satellite. From Fig. 1(a), some redundant features can be observed and the atoms lack high-frequency information. Moreover, some data-specific features can be found in Fig. 1(b). The atoms will contain some objects with larger size of the atom because the atoms are from the raw patches in the HR PAN image. Compared with the atoms in Fig. 1(a) and (b), the filters in Fig. 1(c) show richer variance and diversity in orientation. Besides, we also calculate the incoherence of atoms or filters shown in Fig. 1. The incoherence of atoms from the methods in [24] and [26] are 0.9680 and 0.9303, respectively. The incoherence of filters learned by CSC is 0.7358. As presented in [48], lower incoherence leads to improved approximation performance. Therefore, the filters in Fig. 1(c) can provide a more efficient representation.

Motivated by the advantages of CSC on feature learning and image reconstruction, we proposed a novel pansharpening method based on convolution structure sparse coding (CSSC) by introducing the structure property among the bands of the MS image. First, LR MS and HR PAN images are regarded as the degradation results of the HR MS image in spatial and spectral domains, respectively. The relationship is combined with CSC, and the pansharpening task is converted into a restoration problem. Second, we employ structure sparse constraint to form CSSC model in order to further enhance the correlation and similarity in the bands of MS image. Then, under HR/LR filters, the fusion model is resolved by the alternating direction method of multipliers (ADMM) [36]

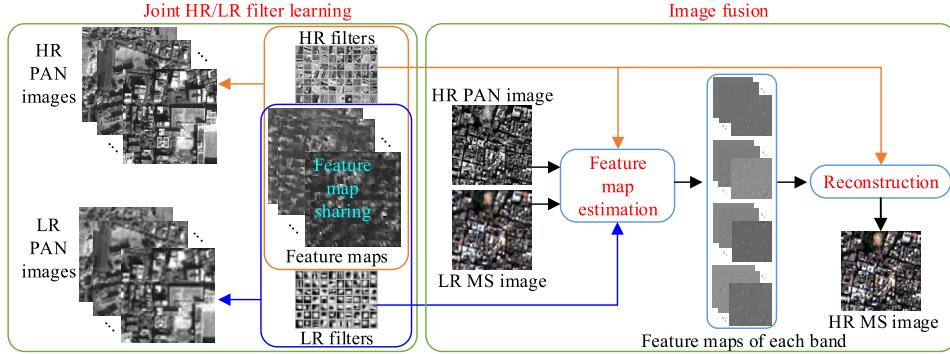


Fig. 2. Flowchart of the proposed image fusion framework. In the framework, the fusion of LR MS and HR PAN images is achieved by the right part. It is assumed that HR and LR filters are known in the fusion model. Therefore, a joint HR/LR filter learning algorithm is developed in the left part.

to obtain the feature maps. Finally, the HR MS image is reconstructed from feature maps via convolution with HR filters. Besides, a joint HR/LR filter learning approach is advanced via block coordinate descent (BCD) method [39] to achieve feature map inferring and filter learning simultaneously. In order to produce compact HR/LR filters, the feature maps of the HR PAN image are shared with those of the LR PAN image. Then, the correspondence among them can be naturally ensured. The flowchart of the proposed method is shown in Fig. 2 to display the whole image fusion process. Finally, the proposed method is tested on different data sets. Visual and numerical performance behaves better than the other compared methods. Meanwhile, we also analyze the influence of filter parameters and regularization parameters on fusion results to provide a more comprehensive investigation. Compared with [33] and [35], the proposed method in this paper employs structure sparse constraint on feature maps to establish the CSSC model with the consideration of band correlation in MS images. To the best of our knowledge, this is the first time that CSC is used for PAN and MS image fusion. Besides, the contributions of this paper are listed as the following two points.

- 1) The LR MS and HR PAN image fusion problem is solved by combining the spatial and spectral degradation models with CSSC. In CSSC,  $L_1$ -norm sparse constraint is replaced by structure sparse prior to capture the high-band correlation among MS.
- 2) A joint HR/LR filter learning framework based on BCD [39] is constructed to ensure the compatibility and correspondence in HR/LR filter pairs. In the optimization algorithm, feature maps and HR/LR filters are updated alternately.

The remainder of this paper is organized as follows. In Section II, the image fusion model based on CSSC and its optimization algorithm are described in detail. In Section III, we establish a joint filter learning model and derive the BCD algorithm for the model. Experimental results and comparisons are presented in Section IV, and conclusions are given in Section V.

## II. IMAGE FUSION BASED ON CONVOLUTION STRUCTURE SPARSE CODING

### A. Convolution Sparse Coding

Traditionally, CSC proposed in [30] models image  $\mathbf{E}$  as the sum of a set of convolutions of the feature maps  $\{\mathbf{Z}_k\}_{k=1,2,\dots,K}$

with their corresponding filters  $\{f_k\}_{k=1,2,\dots,K}$ .  $K$  is the number of feature maps. In order to ensure the sparsity of feature maps,  $L_1$  regularization term is imposed on  $\mathbf{Z}_k$ . Then, the problem is defined as

$$\begin{aligned} \arg \min_{\mathbf{Z}, f} \frac{1}{2} \left\| \mathbf{E} - \sum_{k=1}^K f_k * \mathbf{Z}_k \right\|_2^2 + \lambda \sum_{k=1}^K \|\mathbf{Z}_k\|_1 \\ \text{s.t. } \|f_k\|_2 = 1 \quad \forall k \in \{1, 2, \dots, K\} \end{aligned} \quad (1)$$

where  $\lambda$  is a regularization parameter and  $*$  is the convolution operation.  $\|\cdot\|_1$  is the sum of the absolute values of all elements in  $\mathbf{Z}_k$ . In this formulation, we denote vectors as boldface lowercase letters, and matrices as boldface capitals. The learned filters show less redundancy, and feature maps can well preserve the spatial structures in images. However, it is difficult to optimize because of the convolution operation and the nonconvexity of (1). Besides, the great demands on the computational expense make its application restricted. Zeiler *et al.* [30] solved the problem by alternation optimization algorithm until convergence. Bristow *et al.* [31] proposed an efficient algorithm based on ADMM, in which the subproblem involved convolution is calculated in the Fourier domain. Besides, Wohlberg [32] adopted Sherman–Morrison (SM) [32] to deal with the large linear system in the subproblem of convolution, which reduces the computational complexity a lot and improves the reconstruction performance.

### B. Convolution Structure Sparse Coding for Image Fusion

In our method,  $\mathbf{P} \in R^{\tilde{M} \times \tilde{N}}$  and  $\tilde{\mathbf{M}} \in R^{(\tilde{M}/r) \times (\tilde{N}/r)} \times C$  denote HR PAN and LR MS images, respectively. The desired HR MS image is denoted by  $\mathbf{H} \in R^{\tilde{M} \times \tilde{N} \times C}$ , where  $\tilde{M}$  and  $\tilde{N}$  are the horizontal and vertical sizes of HR PAN image.  $C$  is the number of bands in MS image and  $r$  is the spatial resolution ratio between LR MS and HR PAN images. Generally, the HR PAN image is approximately considered as the linear combination of the bands in the HR MS image. Besides, the LR MS image is regarded as the spatial degraded version of the HR MS image. Thus, relationships of HR PAN and LR MS images between HR MS image can be formulated as

$$\begin{aligned} \mathbf{P} &= \sum_{c=1}^C w_c \mathbf{H}_c + n_1 \\ \mathbf{M}_c &= b_c * \mathbf{H}_c + n_2 \quad c = 1, 2, \dots, C \end{aligned} \quad (2)$$

where  $n_1$  and  $n_2$  are the additive zero-mean random noises.  $w_c$  corresponds the weight on different bands of HR MS image.  $b_c$  is the operator of the point spread function (PSF) for the  $c$ th band of the HR MS image.  $\mathbf{M} \in R^{\tilde{M} \times \tilde{N} \times C}$  is the resampled LR MS image. So,  $\mathbf{M}$  can be seen as the result of the HR MS image after blur operation.

For HR MS image  $\mathbf{H}$ , we assume that it can be decomposed into a series of feature maps with sparse constraint by a set of filters, which is written as

$$\arg \min_{\mathbf{Z}} \frac{1}{2} \sum_{c=1}^C \left\| \mathbf{H}_c - \sum_{k=1}^K f_k * \mathbf{Z}_{k,c} \right\|_2^2 + \lambda \sum_{c=1}^C \sum_{k=1}^K \|\mathbf{Z}_{k,c}\|_1. \quad (3)$$

As we all know, there is a strong structural correlation in the bands of the MS image because of the same observed scene. Naturally, the feature maps corresponding to the same filter for different bands are highly correlated. Therefore, the structure sparsity is imposed on the feature maps to capture the correlation for the CSSC model. Besides, local contrast normalization (LCN) [37], such as removing mean and dividing standard deviations, is generally utilized to produce various discriminative features. However, it is inappropriate to directly use LCN for the image restoration task, such as image fusion, due to its irreversibility. In order to deal with this issue, we introduce a smoothing term  $\mathbf{T}_c$  for each band of HR MS image by adding the  $L_2$ -norm total variation constraint. Then, we can achieve the decomposition of HR MS image with smooth constraint based on CSSC by

$$\begin{aligned} \arg \min_{\mathbf{Z}, \mathbf{T}} & \frac{1}{2} \sum_{c=1}^C \left\| \mathbf{H}_c - \left( \sum_{k=1}^K f_k * \mathbf{Z}_{k,c} + \mathbf{T}_c \right) \right\|_2^2 \\ & + \lambda \sum_{k=1}^K \|\{\mathbf{Z}_{k,c}\}_{c=1,2,\dots,C}\|_{2,1} + \frac{\gamma}{2} \sum_{c=1}^C \|\nabla \mathbf{T}_c\|_2 \end{aligned} \quad (4)$$

where  $\|\cdot\|_{2,1}$  is the  $L_{1,2}$ -norm. Thus,  $\|\{\mathbf{Z}_{k,c}\}_{c=1,2,\dots,C}\|_{2,1}$  is equal  $\sum_{i=1}^{\tilde{M}} \sum_{j=1}^{\tilde{N}} (\sum_{c=1}^C \mathbf{Z}_{k,c}(i, j)^2)^{1/2}$ , which encourages the sparsity in each feature map while minimizing the energy of the corresponding position for all feature maps.  $\nabla$  denotes the difference operator. By introducing the quadratic constraint on  $\nabla \mathbf{T}_c$ , the last term in (4) can ensure the smoothness of  $\mathbf{T}_c$ .  $\gamma$  is a tradeoff parameter. In this way, the feature maps and filters can be efficiently estimated. Besides, smoothing component  $\mathbf{T}_c$  can be learned automatically, which is regarded as the low frequency of each band. Therefore, the spectral information in the HR MS image can be preserved efficiently by structure sparse regularization and smoothing term. Further, for one band in HR MS image,  $\mathbf{T}_c$  can also be regarded as the  $(K+1)$ th feature map  $\mathbf{Z}_{K+1,c}$  with smooth constraint, whose corresponding  $(K+1)$ th filter  $f_{K+1}$  is a 2-D Dirac delta function. Because the convolution result of Dirac delta function with  $\mathbf{T}_c$  is  $\mathbf{T}_c$  itself when ignoring the boundary effect of convolution operation [46]. Thus, the final version about the

decomposition of HR MS image is

$$\begin{aligned} \arg \min_{\mathbf{Z}} & \frac{1}{2} \sum_{c=1}^C \left\| \mathbf{H}_c - \sum_{k=1}^{K+1} f_k * \mathbf{Z}_{k,c} \right\|_2^2 \\ & + \lambda \sum_{k=1}^K \|\{\mathbf{Z}_{k,c}\}_{c=1,2,\dots,C}\|_{2,1} + \frac{\gamma}{2} \sum_{c=1}^C \|\nabla \mathbf{Z}_{K+1,c}\|_2. \end{aligned} \quad (5)$$

Then, combining (2) and (5), we can reformulate pansharpening problem as

$$\begin{aligned} \arg \min_{\mathbf{Z}} & \frac{1}{2} \sum_{c=1}^C \left\| \mathbf{M}_c - b_c * \left( \sum_{k=1}^{K+1} f_k * \mathbf{Z}_{k,c} \right) \right\|_2^2 \\ & + \frac{\alpha}{2} \left\| \mathbf{P} - \sum_{c=1}^C w_c \sum_{k=1}^{K+1} f_k * \mathbf{Z}_{k,c} \right\|_2^2 \\ & + \lambda \sum_{k=1}^K \|\{\mathbf{Z}_{k,c}\}_{c=1,2,\dots,C}\|_{2,1} + \frac{\gamma}{2} \sum_{c=1}^C \|\nabla \mathbf{Z}_{K+1,c}\|_2. \end{aligned} \quad (6)$$

Considering the associativity of convolution and the periodic extension in the Fourier domain of filter, we directly combine  $b_c$  and  $f_k$  to produce LR filters  $d_{k,c}$ . For convenience, the PSF  $b_c$  corresponding to different bands are replaced by the PSF of PAN image in our method to produce LR filters  $d_k$  and simplify the fusion model. So, LR filters  $d_k$  can be learned from the LR PAN images. Then, the improved fusion model can be rewritten as

$$\begin{aligned} \arg \min_{\mathbf{Z}} & \frac{1}{2} \sum_{c=1}^C \left\| \mathbf{M}_c - \sum_{k=1}^{K+1} d_k * \mathbf{Z}_{k,c} \right\|_2^2 \\ & + \frac{\alpha}{2} \left\| \mathbf{P} - \sum_{c=1}^C w_c \sum_{k=1}^{K+1} f_k * \mathbf{Z}_{k,c} \right\|_2^2 \\ & + \lambda \sum_{k=1}^K \|\{\mathbf{Z}_{k,c}\}_{c=1,2,\dots,C}\|_{2,1} + \frac{\gamma}{2} \sum_{c=1}^C \|\nabla \mathbf{Z}_{K+1,c}\|_2. \end{aligned} \quad (7)$$

Finally, HR MS image can be obtained by solving the objective function about CSSC in (7) when  $f_k$  and  $d_k$  are known. The details of the training of HR filters  $f_k$  and LR filters  $d_k$  are elaborated in Section III.

### C. Optimization Algorithm for CSSC

It is difficult to optimize (7) because of the coupling terms containing  $\mathbf{Z}_{k,c}$  and the convolution operation. Thus, the optimization algorithm based on ADMM is derived by introducing auxiliary variables in this section. First, auxiliary variables are introduced in the following equation to decouple

the terms

$$\begin{aligned} & \arg \min_{\mathbf{Z}, \mathbf{X}, \mathbf{U}} \frac{1}{2} \sum_{c=1}^C \left\| \mathbf{M}_c - \sum_{k=1}^{K+1} d_k * \mathbf{X}_{k,c} \right\|_2^2 \\ & + \frac{\alpha}{2} \left\| \mathbf{P} - \sum_{c=1}^C w_c \sum_{k=1}^{K+1} f_k * \mathbf{U}_{k,c} \right\|_2^2 \\ & + \lambda \sum_{k=1}^K \left\| \{\mathbf{Z}_{k,c}\}_{c=1,2,\dots,C} \right\|_{2,1} + \frac{\gamma}{2} \sum_{c=1}^C \left\| \nabla \mathbf{Z}_{K+1,c} \right\|_2 \\ \text{s.t. } & \mathbf{X}_{k,c} = \mathbf{Z}_{k,c}, \quad \mathbf{U}_{k,c} = \mathbf{Z}_{k,c} \\ & k = 1, 2, \dots, K+1, \quad c = 1, 2, \dots, C. \end{aligned} \quad (8)$$

Then, the augmented Lagrange function can be written as follows:

$$\begin{aligned} & \arg \min_{\mathbf{Z}, \mathbf{X}, \mathbf{U}} \frac{1}{2} \sum_{c=1}^C \left\| \mathbf{M}_c - \sum_{k=1}^{K+1} d_k * \mathbf{X}_{k,c} \right\|_2^2 \\ & + \frac{\alpha}{2} \left\| \mathbf{P} - \sum_{c=1}^C w_c \sum_{k=1}^{K+1} f_k * \mathbf{U}_{k,c} \right\|_2^2 \\ & + \lambda \sum_{k=1}^K \left\| \{\mathbf{Z}_{k,c}\}_{c=1,2,\dots,C} \right\|_{2,1} + \frac{\gamma}{2} \sum_{c=1}^C \left\| \nabla \mathbf{Z}_{K+1,c} \right\|_2 \\ & + \sum_{c=1}^C \sum_{k=1}^{K+1} \left( \frac{\mu_1}{2} \|\mathbf{X}_{k,c} - \mathbf{Z}_{k,c}\|_2^2 + \langle \mathbf{Y}_{k,c}, \mathbf{X}_{k,c} - \mathbf{Z}_{k,c} \rangle \right) \\ & + \sum_{c=1}^C \sum_{k=1}^{K+1} \left( \frac{\mu_2}{2} \|\mathbf{U}_{k,c} - \mathbf{Z}_{k,c}\|_2^2 + \langle \mathbf{V}_{k,c}, \mathbf{U}_{k,c} - \mathbf{Z}_{k,c} \rangle \right). \end{aligned} \quad (9)$$

So, under the framework of ADMM,  $\mathbf{X}_{k,c}$ ,  $\mathbf{U}_{k,c}$ , and  $\mathbf{Z}_{k,c}$  are solved by alternative optimization when the other variables are fixed.

### 1) Subproblem for $\mathbf{X}_{k,c}$ :

$$\begin{aligned} & \arg \min_{\mathbf{X}} \frac{1}{2} \sum_{c=1}^C \left\| \mathbf{M}_c - \sum_{k=1}^{K+1} d_k * \mathbf{X}_{k,c} \right\|_2^2 \\ & + \sum_{c=1}^C \sum_{k=1}^{K+1} \frac{\mu_1}{2} \|\mathbf{X}_{k,c} - \mathbf{Q}_{k,c}\|_2^2. \end{aligned} \quad (10)$$

Equation (10) is the subproblem about  $\mathbf{X}_{k,c}$ , where  $\mathbf{Q}_{k,c}$  is equal to  $\mathbf{Z}_{k,c} - (\langle \mathbf{Y}_{k,c}, \mathbf{X}_{k,c} \rangle / \mu_1)$ . Owing to the independence of each band, feature maps of each band can be solved independently. Besides, because the convolution operation corresponds to the element-wise multiplication in the frequency domain, (10) is solved in the Fourier domain more efficiently. Then, the problem for  $c$ th band can be expressed as

$$\arg \min_{\hat{\mathbf{x}}} \frac{1}{2} \left\| \hat{\mathbf{m}}_c - \sum_{k=1}^{K+1} \hat{\mathbf{D}}_k \hat{\mathbf{x}}_{k,c} \right\|_2^2 + \sum_{k=1}^{K+1} \frac{\mu_1}{2} \|\hat{\mathbf{x}}_{k,c} - \hat{\mathbf{q}}_{k,c}\|_2^2. \quad (11)$$

$\hat{\cdot}$  denotes the result of discrete Fourier transform (DFT). So,  $\hat{\mathbf{m}}_c$ ,  $\hat{\mathbf{x}}_{k,c}$ , and  $\hat{\mathbf{q}}_{k,c}$  are the vector form of  $\mathbf{M}_c$ ,  $\mathbf{X}_{k,c}$ , and  $\mathbf{Q}_{k,c}$  after DFT, respectively.  $\hat{\mathbf{D}}_k$  is a diagonal matrix whose

diagonal elements are the results of  $d_k$  in the DFT domain. For more concise from, (11) is rewritten as

$$\arg \min_{\hat{\mathbf{x}}} \frac{1}{2} \|\hat{\mathbf{m}}_c - \hat{\mathbf{D}} \hat{\mathbf{x}}_c\|_2^2 + \frac{\mu_1}{2} \|\hat{\mathbf{x}}_c - \hat{\mathbf{q}}_c\|_2^2 \quad (12)$$

where  $\hat{\mathbf{D}} = [\hat{\mathbf{D}}_1, \hat{\mathbf{D}}_2, \dots, \hat{\mathbf{D}}_K]$ ,  $\hat{\mathbf{x}}_c = [\hat{\mathbf{x}}_{1,c}^T, \hat{\mathbf{x}}_{2,c}^T, \dots, \hat{\mathbf{x}}_{K,c}^T]^T$ , and  $\hat{\mathbf{q}}_c = [\hat{\mathbf{q}}_{1,c}^T, \hat{\mathbf{q}}_{2,c}^T, \dots, \hat{\mathbf{q}}_{K,c}^T]^T$ . A closed-form solution of (12) can be derived by solving  $D$  independent  $K \times K$  linear systems in the following equation.  $D$  is the dimension of  $\hat{\mathbf{m}}_c$

$$(\hat{\mathbf{D}}^H \hat{\mathbf{D}} + \mu_1 \mathbf{I}) \hat{\mathbf{x}}_c = \hat{\mathbf{D}}^H \hat{\mathbf{m}}_c + \mu_1 \hat{\mathbf{q}}_c \quad (13)$$

where  $\mathbf{I}$  is the identity matrix with matched size.  $H$  is the Hermitian transpose. Conjugate gradient method is adopted in [31] to solve the above linear system. In order to further decrease the complexity, SM [32] is considered in our proposed method to achieve faster computation with acceptable accuracy.

### 2) Subproblem for $\mathbf{U}_{k,c}$ :

$$\begin{aligned} & \arg \min_{\mathbf{U}} \frac{\alpha}{2} \left\| \mathbf{P} - \sum_{c=1}^C w_c \sum_{k=1}^{K+1} f_k * \mathbf{U}_{k,c} \right\|_2^2 \\ & + \sum_{c=1}^C \sum_{k=1}^{K+1} \frac{\mu_2}{2} \|\mathbf{U}_{k,c} - \mathbf{O}_{k,c}\|_2^2 \end{aligned} \quad (14)$$

where  $\mathbf{O}_{k,c}$  is  $\mathbf{Z}_{k,c} - ((\mathbf{V}_{k,c}) / \mu_2)$ . The subproblem of  $\mathbf{U}_{k,c}$  in (14) can be rearranged as (15) in the same way to (11). Then, (14) is equivalent to

$$\arg \min_{\hat{\mathbf{u}}} \frac{\alpha}{2} \|\hat{\mathbf{p}} - \hat{\mathbf{F}}_A \hat{\mathbf{u}}\|_2^2 + \frac{\mu_2}{2} \|\hat{\mathbf{u}} - \hat{\mathbf{o}}\|_2^2 \quad (15)$$

where  $\hat{\mathbf{F}}_A$  equals to  $[\hat{\mathbf{F}}_1, \hat{\mathbf{F}}_2, \dots, \hat{\mathbf{F}}_C]$  and  $\hat{\mathbf{F}}_c = w_c [\text{diag}(\text{vec}(\hat{f}_1)), \text{diag}(\text{vec}(\hat{f}_2)), \dots, \text{diag}(\text{vec}(\hat{f}_K))]$ .  $\text{diag}$  transforms a vector into a diagonal matrix and  $\text{vec}$  is the vectorization operation.  $\hat{\mathbf{p}}$  is equal to  $\text{vec}(\hat{\mathbf{P}})$ .  $\hat{\mathbf{u}}$  and  $\hat{\mathbf{o}}$  are the concatenation of all  $\mathbf{U}_{k,c}$  and  $\mathbf{O}_{k,c}$  in the vector form, respectively. Then, the solution of (14) is given in the following equation:

$$(\alpha \hat{\mathbf{F}}_A^H \hat{\mathbf{F}}_A + \mu_2 \mathbf{I}) \hat{\mathbf{u}} = \alpha \hat{\mathbf{F}}_A^H \hat{\mathbf{p}} + \mu_2 \hat{\mathbf{o}}. \quad (16)$$

Similarly,  $\mathbf{U}_{k,c}$  is updated to compute (16) through SM [32].

### 3) Subproblem for $\mathbf{Z}_{k,c}$ :

$$\begin{aligned} & \arg \min_{\mathbf{Z}} \lambda \sum_{k=1}^K \left\| \{\mathbf{Z}_{k,c}\}_{c=1,2,\dots,C} \right\|_{2,1} + \frac{\gamma}{2} \sum_{c=1}^C \left\| \nabla \mathbf{Z}_{K+1,c} \right\|_2 \\ & + \sum_{c=1}^C \sum_{k=1}^{K+1} \left( \frac{\mu_1}{2} \left\| \mathbf{Z}_{k,c} - \left( \mathbf{X}_{k,c} + \frac{1}{\mu_1} \mathbf{Y}_{k,c} \right) \right\|_2^2 \right) \\ & + \sum_{c=1}^C \sum_{k=1}^{K+1} \left( \frac{\mu_2}{2} \left\| \mathbf{Z}_{k,c} - \left( \mathbf{U}_{k,c} + \frac{1}{\mu_2} \mathbf{V}_{k,c} \right) \right\|_2^2 \right). \end{aligned} \quad (17)$$

Due to different constraints on  $\mathbf{Z}_{k,c}$ ,  $k = 1, 2, \dots, K$  and  $\mathbf{Z}_{K+1,c}$ , we first optimize  $\mathbf{Z}_{k,c}$ ,  $k = 1, 2, \dots, K$  independently by the following equation [38]:

$$\begin{aligned} & \mathbf{Z}_k(i, j, :) \\ & = \begin{cases} \frac{\|\mathbf{S}_k(i, j, :) - \theta\|_2}{\|\mathbf{S}_k(i, j, :)\|_2} \mathbf{S}_k(i, j, :), & \text{if } \theta < \|\mathbf{S}_k(i, j, :)\|_2 \\ 0, & \text{otherwise} \end{cases} \end{aligned} \quad (18)$$

**Algorithm 1** CSSC Fusion via ADMM

**Input:** LR MS image  $\mathbf{M}$ , HR PAN image  $\mathbf{P}$ , HR filters  $f_k$ , LR filters  $d_k$  PSFs  $b_c$ , spectral weights  $w_c$ , parameters  $\alpha$ ,  $\lambda$  and  $\gamma$ ;

**Initialization:**  $\mathbf{X}$ ,  $\mathbf{U}$  and  $\mathbf{Z}$  are initialized by random values in  $[0, 1]$ ,  $\mathbf{Y} = \mathbf{V} = \mathbf{0}$ ,  $\mu_1 = \mu_2 = 0.5$ ,  $\rho = 1.3$ ,  $\max_{\mu} = 10^{15}$ ;

**While** not converged **do**

1. Update  $\mathbf{X}_{k,c}$ ,  $\forall k, \forall c$  by solving (13) through SM;
2. Update  $\mathbf{U}_{k,c}$ ,  $\forall k, \forall c$  by solving (16) through SM;
3. Update  $\mathbf{Z}_{k,c}$ ,  $k = 1, 2, \dots, K, \forall c$  by (18);  
Update  $\mathbf{Z}_{K+1,c}$ ,  $\forall c$  by (20);
4. Update multipliers  $\mathbf{Y}_{k,c}$  and  $\mathbf{V}_{k,c}$  by (21)
5.  $\mu_1^{t+1} \leftarrow \min(\rho \mu_1^t, \max_{\mu}), \mu_2^{t+1} \leftarrow \min(\rho \mu_2^t, \max_{\mu})$ ;
6.  $t = t + 1$ ;

**end while;**

**Output:** HR MS image  $\mathbf{H}$ .

where  $\mathbf{S}_k$  equals to  $((\mu_1 \mathbf{X}_k + \mathbf{Y}_k + \mu_2 \mathbf{U}_k + \mathbf{V}_k)/(\mu_1 + \mu_2))$  and  $\theta$  is  $(\lambda/(\mu_1 + \mu_2))$ .

For  $\mathbf{Z}_{K+1,c}$ , it can be obtained by solving (19) band by band

$$\begin{aligned} \arg \min_{\mathbf{Z}_{K+1}} & \frac{\gamma}{2} \|\nabla \mathbf{Z}_{K+1,c}\|_2^2 \\ & + \frac{\mu_2}{2} \left\| \mathbf{Z}_{K+1,c} - \left( \mathbf{U}_{K+1,c} + \frac{1}{\mu_2} \mathbf{V}_{K+1,c} \right) \right\|_2^2 \\ & + \frac{\mu_1}{2} \left\| \mathbf{Z}_{K+1,c} - \left( \mathbf{X}_{K+1,c} + \frac{1}{\mu_1} \mathbf{Y}_{K+1,c} \right) \right\|_2^2. \quad (19) \end{aligned}$$

Then, the closed-form solution for this least-squares minimization problem can be calculated by following equation in the Fourier domain:

$$\begin{aligned} \mathbf{Z}_{K+1,c} &= \mathcal{F}^{-1} \left( \frac{\mu_1 \hat{\mathbf{X}}_{K+1,c} + \hat{\mathbf{Y}}_{K+1,c} + \mu_2 \hat{\mathbf{U}}_{K+1,c} + \hat{\mathbf{V}}_{K+1,c}}{\gamma (\mathcal{F}(\nabla_h) \mathcal{F}(\nabla_h) + \mathcal{F}(\nabla_v) \mathcal{F}(\nabla_v)) + \mu_1 + \mu_2} \right) \\ & \quad (20) \end{aligned}$$

where  $\mathcal{F}$  is the Fourier transform and  $\mathcal{F}^{-1}$  is the inverse version of  $\mathcal{F}$ .  $\overline{\mathcal{F}(\cdot)}$  denotes the complex conjugate operator.  $\nabla_h$  and  $\nabla_v$  are the horizontal and vertical differential operators, respectively.

Finally, the Lagrange multipliers can be computed by the following equation:

$$\begin{aligned} \mathbf{Y}_{k,c} &= \mathbf{Y}_{k,c} + \mu_1 (\mathbf{X}_{k,c} - \mathbf{Z}_{k,c}) \\ \mathbf{V}_{k,c} &= \mathbf{V}_{k,c} + \mu_2 (\mathbf{U}_{k,c} - \mathbf{Z}_{k,c}). \quad (21) \end{aligned}$$

Moreover, the penalty parameters  $\mu_1$  and  $\mu_2$  increase dynamically by multiplying a small value. The stop condition of the algorithm is decided by the maximal number of iterations or reconstruction errors of HR PAN and LR MS images. Generally, we set the maximum number of iterations as 400. The update steps are provided in Algorithm 1. Finally, the HR MS image can be reconstructed by the following equation once the feature maps are estimated

$$\mathbf{H}_c = \sum_{k=1}^{K+1} f_k * \mathbf{Z}_{k,c} \quad \forall c \in \{1, 2, \dots, C\}. \quad (22)$$

**D. Complexity Analysis of CSSC**

In this section, we briefly analyze the computational complexity of the proposed image fused method. Here, the cost of a single iteration of each subproblem is considered. One can see that the optimization algorithm for CSSC consists of three subproblems. It involves Fourier transform and the solution to linear systems when update  $\mathbf{X}_{k,c}$ . The complexity of Fourier transform is  $C(K+1)D \log D$ . For the linear systems, SM [32] is applied whose complexity is  $C(K+1)D$ . So, the complexity of updating  $\mathbf{X}_{k,c}$  is  $C(K+1)D \log D + C(K+1)D$ . For the calculation of  $\mathbf{U}_{k,c}$ , the update scheme is also achieved by SM. Then, the complexity of updating  $\mathbf{U}_{k,c}$  is the same as that of  $\mathbf{X}_{k,c}$ . In the subproblem of  $\mathbf{Z}_{k,c}$ , the first  $K$  feature maps and  $(K+1)$ th feature map are computed separately, whose complexities are  $C^2KD + KD$  and  $CD \log D$ , respectively.

**III. JOINT HR/LR FILTER LEARNING**

In the above CSSC-based fusion model, HR PAN and LR MS images are constructed by HR filters and LR filters with the same feature maps, respectively. Due to the sharing strategy of feature maps, compact HR/LR filters have a significant impact on the fused image. Although HR/LR filters can be obtained separately, the correspondence between them cannot be achieved [47]. Therefore, it is vital to learn fine filters while ensuring the correspondence in HR/LR filters [50], [51]. Then, the HR/LR filter learning model is given in this section and the optimization algorithm is also derived via alternating minimization with respect to feature maps and filters. Then, the filter learning framework is modeled as

$$\begin{aligned} \arg \min_{f,d,\mathbf{Z}} & \frac{1}{2} \sum_{n=1}^N \left\| \sum_{k=1}^K f_k * \mathbf{Z}_{k,n} - \mathbf{P}_n \right\|_2^2 + \beta \sum_{n=1}^N \sum_{k=1}^K \|\mathbf{Z}_{k,n}\|_1 \\ & + \frac{1}{2} \sum_{n=1}^N \left\| \sum_{k=1}^K d_k * \mathbf{Z}_{k,n} - \mathbf{L}_n \right\|_2^2 \\ \text{s.t. } & \|f_k\|_2 = 1, \quad \|d_k\|_2 = 1 \quad \forall k \quad (23) \end{aligned}$$

where  $\mathbf{P}_n \in \{\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_N\}$  and  $\mathbf{L}_n \in \{\mathbf{L}_1, \mathbf{L}_2, \dots, \mathbf{L}_N\}$  are the HR/LR PAN image pairs for the training of the HR/LR filter pairs  $f_k$  and  $d_k$ . The LR PAN images are produced from HR PAN images by down-sampling with a factor of 4.  $N$  is the number of training images. The normalization constraint on filters  $f_k$  and  $d_k$  is used to avoid the scaling ambiguity.  $\beta$  controls the sparsity of feature maps. Through this model, the correspondence between  $f_k$  and  $d_k$  is considered by sharing the same feature maps. Considering the multiconvex optimization problem, BCD method [39] is exploited to deal with (23) for stability and good convergence.

**A. Feature Map Inferring**

The subproblem about  $\mathbf{Z}_{k,n}$  can be defined as

$$\begin{aligned} \arg \min_{\mathbf{Z}} & \frac{1}{2} \sum_{n=1}^N \left\| \sum_{k=1}^K f_k * \mathbf{Z}_{k,n} - \mathbf{P}_n \right\|_2^2 + \beta \sum_{n=1}^N \sum_{k=1}^K \|\mathbf{Z}_{k,n}\|_1 \\ & + \frac{1}{2} \sum_{n=1}^N \left\| \sum_{k=1}^K d_k * \mathbf{Z}_{k,n} - \mathbf{L}_n \right\|_2^2. \quad (24) \end{aligned}$$

Naturally, the auxiliary variable  $\mathbf{T}_{k,n}$  associated with  $\mathbf{Z}_{k,n}$  is introduced to decouple the sparse term. Further, the augmented Lagrange function can be written as

$$\begin{aligned} \arg \min_{\mathbf{Z}, \mathbf{T}} & \frac{1}{2} \sum_{n=1}^N \left\| \sum_{k=1}^K f_k * \mathbf{Z}_{k,n} - \mathbf{P}_n \right\|_2^2 + \beta \sum_{n=1}^N \sum_{k=1}^K \|\mathbf{T}_{k,n}\|_1 \\ & + \frac{1}{2} \sum_{n=1}^N \left\| \sum_{k=1}^K d_k * \mathbf{Z}_{k,n} - \mathbf{L}_n \right\|_2^2 \\ & + \sum_{n=1}^N \sum_{k=1}^K \left( \frac{\sigma_1}{2} \|\mathbf{T}_{k,n} - \mathbf{Z}_{k,n}\|_2^2 + \langle \mathbf{G}_{k,n}, \mathbf{T}_{k,n} - \mathbf{Z}_{k,n} \rangle \right) \end{aligned} \quad (25)$$

where  $\sigma_1$  is the penalty parameter and  $\mathbf{G}_{k,n}$  is the Lagrange multiplier. Thus, it can be solved by the ADMM algorithm.

### 1) Subproblem for $\mathbf{Z}_{k,n}$ :

$$\begin{aligned} \arg \min_{\mathbf{Z}} & \frac{1}{2} \sum_{n=1}^N \left\| \sum_{k=1}^K f_k * \mathbf{Z}_{k,n} - \mathbf{P}_n \right\|_2^2 \\ & + \frac{1}{2} \sum_{n=1}^N \left\| \sum_{k=1}^K d_k * \mathbf{Z}_{k,n} - \mathbf{L}_n \right\|_2^2 \\ & + \sum_{n=1}^N \sum_{k=1}^K \frac{\sigma_1}{2} \left\| \mathbf{T}_{k,n} - \mathbf{Z}_{k,n} + \frac{1}{\sigma_1} \mathbf{G}_{k,n} \right\|_2^2. \end{aligned} \quad (26)$$

Then, the feature maps of each image can be computed independently. According to the same way to the subproblem for  $\mathbf{X}_{k,c}$ , (26) is transformed into the Fourier domain to derive the closed-form solution of  $\mathbf{Z}_{k,n}$ . So, the minimization with respect to  $\mathbf{Z}_{k,n}$  can be reformulated as

$$\begin{aligned} \arg \min_{\hat{\mathbf{z}}_n} & \frac{1}{2} \|\hat{\mathbf{F}} \hat{\mathbf{z}}_n - \hat{\mathbf{p}}_n\|_2^2 + \frac{1}{2} \|\hat{\mathbf{D}} \hat{\mathbf{z}}_n - \hat{\mathbf{l}}_n\|_2^2 \\ & + \frac{\sigma_1}{2} \left\| \hat{\mathbf{z}}_n - \hat{\mathbf{t}}_n - \frac{1}{\sigma_1} \hat{\mathbf{g}}_n \right\|_2^2 \end{aligned} \quad (27)$$

where  $\hat{\mathbf{F}} = [\text{diag}(\text{vec}(\hat{f}_1)), \text{diag}(\text{vec}(\hat{f}_2)), \dots, \text{diag}(\text{vec}(\hat{f}_K))]$  and  $\hat{\mathbf{D}}$  follows the same arrangement.  $\hat{\mathbf{l}}_n$  is the vector form of  $\mathbf{L}_n$  after DFT.  $\hat{\mathbf{t}}_n$  and  $\hat{\mathbf{g}}_n$  are the concatenation of all  $\mathbf{T}_{k,n}$  and  $\mathbf{G}_{k,n}$  in vector after DFT, respectively. By calculating the derivation of (27) with respect to  $\hat{\mathbf{z}}_n$  and making the first partial derivatives equal to 0,  $\hat{\mathbf{z}}_n$  can be obtained through iterated SM [32].

### 2) Subproblem for $\mathbf{T}_{k,n}$ :

$$\beta \sum_{n=1}^N \sum_{k=1}^K \|\mathbf{T}_{k,n}\|_1 + \sum_{n=1}^N \sum_{k=1}^K \frac{\sigma_1}{2} \left\| \mathbf{T}_{k,n} - \mathbf{Z}_{k,n} + \frac{1}{\sigma_1} \mathbf{G}_{k,n} \right\|_2^2. \quad (28)$$

Obviously, it can be efficiently computed by the soft-thresholding shrinkage operator [32] according to the following equation:

$$\mathbf{T}_{k,n} = \mathcal{S}_{\beta/\sigma_1} \left( \mathbf{Z}_{k,n} - \frac{1}{\sigma_1} \mathbf{G}_{k,n} \right) \quad (29)$$

where  $\mathcal{S}_v(a) = \text{sign}(a) \odot \max(0, |a| - v)$  and  $v$  is  $\beta/\sigma_1$ .  $\odot$  denotes element-wise multiplication.

For the Lagrange multiplier  $\mathbf{G}_{k,n}$ , it is updated according to the rule in ADMM. Besides,  $\sigma_1$  becomes larger with the

increasing of iteration. Finally, the inferring of feature maps can be achieved by updating  $\mathbf{Z}_{k,n}$  and  $\mathbf{T}_{k,n}$  alternatively.

### B. Filter Learning

There are two kinds of filters to be trained in the framework. Here, the HR filters  $f_k$  are optimized first. Considering the consistency of spatial support, an auxiliary variable  $r_k$  is introduced for  $f_k$  to optimize the problem in the frequency domain and spatial domain, respectively. Thus, the objective of HR filters  $f_k$  can be defined as

$$\arg \min_{\hat{\mathbf{F}}, r} \frac{1}{2} \sum_{n=1}^N \left\| \sum_{k=1}^K \hat{\mathbf{F}}_k \odot \hat{\mathbf{Z}}_{k,n} - \hat{\mathbf{p}}_n \right\|_2^2$$

$$\text{s.t. } r_k = \mathcal{F}^{-1}(\hat{\mathbf{F}}_k), \quad r_k = P(r_k), \quad \|r_k\|_2 = 1 \quad \forall k \quad (30)$$

where  $P(\cdot)$  is a mapping operator, which makes the coefficients outside the spatial support to be 0.  $\hat{\mathbf{Z}}_{k,n}$  and  $\hat{\mathbf{F}}_k$  are the DFT versions of  $\mathbf{Z}_{k,n}$  and  $f_k$ , respectively. Then, the augmented Lagrange function of (30) is

$$\begin{aligned} \arg \min_{\hat{\mathbf{F}}, \mathbf{r}} & \frac{1}{2} \sum_{n=1}^N \|\hat{\mathbf{Z}}_n \hat{\mathbf{F}} - \hat{\mathbf{p}}_n\|_2^2 \\ & + \sum_{k=1}^K \left( \frac{\sigma_2}{2} \|r_k - \mathcal{F}^{-1}(\hat{\mathbf{F}}_k)\|_2^2 + \langle \mathbf{j}_k, r_k - \mathcal{F}^{-1}(\hat{\mathbf{F}}_k) \rangle \right) \end{aligned}$$

$$\text{s.t. } \mathbf{r}_k = P(\mathbf{r}_k), \quad \|\mathbf{r}_k\|_2 = 1 \quad \forall k \quad (31)$$

where  $\hat{\mathbf{f}}$  is the concatenated result of all vectorized  $\hat{\mathbf{F}}_k$  and  $\hat{\mathbf{Z}}_n$  is  $[\text{diag}(\text{vec}(\mathbf{Z}_{1,n})), \text{diag}(\text{vec}(\mathbf{Z}_{2,n})), \dots, \text{diag}(\text{vec}(\mathbf{Z}_{K,n}))]$ .  $\mathbf{r}_k$  stands for the vectorized result of  $r_k$ . Moreover,  $\mathbf{j}_k$  and  $\sigma_2$  are the Langrange multiplier and the penalty parameter, respectively. Then, it is solved alternatively as follows.

*1) Subproblem for  $f_k$ :* In order to obtain more elaborate filters, all images are exploited to update them simultaneously. Besides, omitting the inverse Fourier transform in (31), the more concise subproblem can be rewritten as

$$\arg \min_{\hat{\mathbf{F}}} \frac{1}{2} \sum_{n=1}^N \|\hat{\mathbf{Z}}_n \hat{\mathbf{F}} - \hat{\mathbf{p}}_n\|_2^2 + \frac{\sigma_2}{2} \|\hat{\mathbf{r}} - \hat{\mathbf{f}}\|_2^2 + \langle \hat{\mathbf{j}}, \hat{\mathbf{f}} - \hat{\mathbf{r}} \rangle \quad (32)$$

where  $\hat{\mathbf{r}} = [\hat{\mathbf{r}}_1^T, \hat{\mathbf{r}}_2^T, \dots, \hat{\mathbf{r}}_K^T]^T$  and  $\hat{\mathbf{j}} = [\hat{\mathbf{j}}_1^T, \hat{\mathbf{j}}_2^T, \dots, \hat{\mathbf{j}}_K^T]^T$ . Then, the solution of (32) also can be computed in the same way as (27) by using iterated SM.

### 2) Subproblem for $r_k$ :

$$\begin{aligned} \arg \min_{\mathbf{r}} & \frac{\sigma_2}{2} \|\mathbf{r}_k - \mathcal{F}^{-1}(\hat{\mathbf{F}}_k)\|_2^2 + \langle \mathbf{j}_k, \mathbf{r}_k - \mathcal{F}^{-1}(\hat{\mathbf{F}}_k) \rangle \\ \text{s.t. } & \mathbf{r}_k = P(\mathbf{r}_k), \quad \|\mathbf{r}_k\|_2 = 1. \end{aligned} \quad (33)$$

Thus, (33) is simplified as follows by introducing the spatial support constraint:

$$\begin{aligned} \arg \min_{\mathbf{r}} & \frac{\sigma_2}{2} \|\mathbf{r}_k - P(\mathcal{F}^{-1}(\hat{\mathbf{F}}_k))\|_2^2 + \langle P(\mathbf{j}_k), \mathbf{r}_k - P(\mathcal{F}^{-1}(\hat{\mathbf{F}}_k)) \rangle \\ \text{s.t. } & \|\mathbf{r}_k\|_2 = 1. \end{aligned} \quad (34)$$

Then, the solution to the unconstrained version of (34) can be computed through

$$\mathbf{r}_k = P(\mathcal{F}^{-1}(\hat{\mathbf{F}}_k)) - \frac{1}{\sigma_2} P(\mathbf{j}_k). \quad (35)$$

Finally, the normalization is implemented for the estimated  $\mathbf{r}_k$ .

Besides,  $\mathbf{j}_k$  is calculated by the update rule. For the subproblem for  $d_k$ , it is similar to that of  $f_k$ . Thus,  $d_k$  is updated in the same way to  $f_k$ . Besides, we give the flowchart of the optimization algorithm based on BCD in Algorithm 2. We set the number of outer iterations  $T_1$  as 1000 to produce finer filters.  $J_1$ ,  $J_2$  and  $J_3$  are set as 10.

#### Algorithm 2 Filter Learning via BCD

**Input:** HR PAN image set  $\{\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_N\}$ , LR PAN image set  $\{\mathbf{L}_1, \mathbf{L}_2, \dots, \mathbf{L}_N\}$  parameter  $\beta$ ;

**Initialization:**  $f$ ,  $d$  and  $\mathbf{Z}$  are initialized by random values in  $[0, 1]$ ,

**Outer loop:** for  $t_1 = 1, 2, \dots, T_1$  do

**Feature Map Inferring:** for  $j_1 = 1, 2, \dots, J_1$  do

- (1) Update  $\mathbf{Z}_{k,n}, \forall k, \forall n$  by solving (27) through iterated SM;
- (2) Update  $\mathbf{T}_{k,n}, \forall k, \forall n$  by (29);

- (3) Update multiplier  $\mathbf{G}_{k,n}, \forall k, \forall n$ ;

- (4) Output  $\mathbf{Z}_{k,n}$  if  $j_1 = J_1$ .

**End for**

**HR filter update:** for  $j_2 = 1, 2, \dots, J_2$  do

- (1) Update  $f_k, \forall k$  by solving (32) through iterated SM;

- (2) Update  $r_k, \forall k$  by (35);

- (3) Update multiplier  $\mathbf{j}_k$ ;

- (4) Output  $f_k$  if  $j_2 = J_2$ .

**End for**

**LR filter update:** for  $j_3 = 1, 2, \dots, J_3$  do

        Learning LR filter in the same way to update HR filters

**End for**

**Output:** HR filter  $f_k$  and LR filter  $d_k$  if  $t_1 = T_1$ .

**End for**

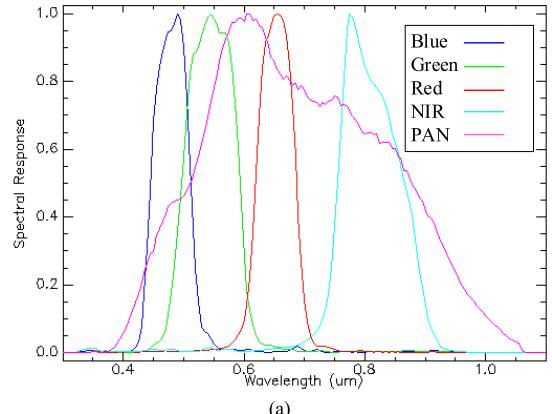
#### C. Complexity Analysis of HR/LR Filter Learning

In the joint HR/LR learning algorithm, we alternately calculate feature maps, HR and LR filters. When infer feature maps, the maps of each training image are updated sequentially. Thus, taking Fourier transform and the iterative SM [32] into consideration, the complexity of computing  $\mathbf{Z}_{k,n}$  of all images is  $NKD \log D + 2NKD$ . For the auxiliary variable  $\mathbf{T}_{k,n}$ , the complexity is  $NKD$  due to using soft-thresholding shrinkage. In the second stage, all training images are simultaneously utilized to estimate HR filters. For the subproblem of  $f_k$ , Fourier transform and the iterative SM [32] are implemented. Then, the complexity is  $KD \log D + NKD$ . The complexity of updating  $r_k$  is  $KD \log D$ . Because LR filters are optimized in the same way to  $f_k$ , the complexity of  $d_k$  is the same as that of HR filters. As described above, we can see that the complexity of the whole training stage is dominated by the estimation of feature maps.

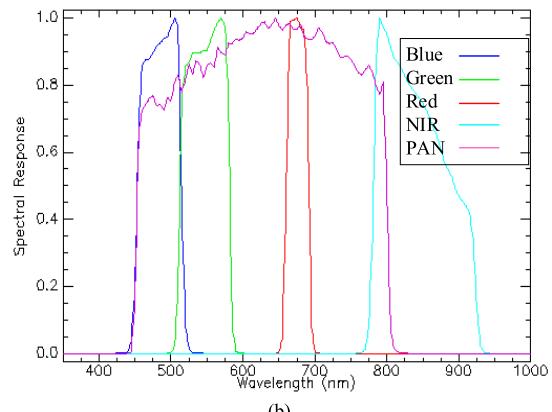
## IV. EXPERIMENTAL RESULTS AND COMPARISONS

### A. Experiment Setup

In this section, several classical methods are compared with the proposed methods, such as GS [7], GIHS [10], AWLP [12], SVT [16], SRLD [24], and SparseFI [26]. Besides, comparison



(a)



(b)

Fig. 3. Spectral response function for different satellites. (a) QuickBird. (b) Geoeye-1.

is conducted on reduced-scale data set and full-scale data set from QuickBird and Geoeye-1 satellites for a comprehensive analysis. The spectral response functions of QuickBird and Geoeye-1 satellites are shown in Fig. 3, respectively. Reduced-scale means that the original MS and PAN images are degraded by modulation transfer function [40] and decimation. The down-sampling rate is 4. Then, the original MS image is regarded as the reference image to compare with the fusion result of degraded source images. In full-scale data set, the original MS and PAN images are directly fused to produce the HR MS image whose resolution is the same as that of PAN image. For QuickBird, the data are acquired from Sundarbans, India, on November 21, 2002, which can provide the four-band MS image at 2.8-m spatial resolution and the PAN image at 0.7-m spatial resolution. The data from Geoeye-1 cover the urban area of Hobart, Australia, collected on February 24, 2009. The spatial resolution of PAN image from the data is 0.5 m, and the resolution of MS image is 2 m. The spatial sizes of PAN image and MS image to be fused are  $256 \times 256$  and  $64 \times 64$ , respectively.

For the reduced-scale data set, Q4 [41], spectral angle mapper (SAM) [42], universal image quality index (UIQI) [43], and *Erreur Relative Globale Adimensionnelle de Synthèse* (ERGAS) [44] are employed to evaluate the quality of fusion results according to the Wald's protocol [44]. Generally, larger Q4 and UIQI mean better fusion results. For SAM and ERGAS, smaller values denote better results and the optimal values are 0. For full-scale data set, quality no



Fig. 4. Fusion results of images from QuickBird. (a) Resampled LR MS image (b) HR PAN image. (c) Reference HR MS image. (d) GIHS [10]. (e) GS [7]. (f) AWLP [12]. (g) SVT [16]. (h) SRLD [24]. (i) SparseFI [26]. (j) CSSC.

reference (QNR) [45] is calculated by  $D_S$  [45] and  $D_\lambda$  [45].  $D_S$  and  $D_\lambda$  are employed to measure the spatial and spectral distortion, respectively. The larger the QNR, the better is the fused image. For  $D_S$  and  $D_\lambda$ , the best values are 0.

In the proposed method, the number of feature maps  $K$  is equal to 70 and the size of filter is  $9 \times 9$ . For the data from QuickBird, parameter  $\alpha$  of PAN image fidelity is set as  $2^3$ . The tradeoff parameters about feature maps  $\lambda$  and  $\gamma$  are set as 4 and 4, respectively.  $\alpha$ ,  $\lambda$ , and  $\gamma$  are set as  $2^4$ , 4, and 1 for the data from Geoeye-1. In the HR/LR filter learning part,  $\beta$  is equal to 0.2. The weight settings in (1) for QuickBird satellite [22] are:  $w_1 = 0.1139$ ,  $w_2 = 0.2315$ ,  $w_3 = 0.2308$ , and  $w_4 = 0.4239$ . For Geoeye-1, the weights in (1) are set as [49]:  $w_1 = 0.3168$ ,  $w_2 = 0.3787$ ,  $w_3 = 0.1964$ , and  $w_4 = 0.1081$ .

### B. Training Data Sets

In Section IV, comparison experiments are conducted on reduced-scale and full-scale data sets from QuickBird and Geoeye-1 satellites. Thus, four sets of HR/LR filters should be trained for different data sets. For reduced-scale training data set about QuickBird, we randomly select 50 PAN images with a size of  $1024 \times 1024$  captured from Sundarbans, India, and then PAN images are degraded spatially with down-sampling by a factor of four to produce the reduced-scale PAN images for filter learning. Besides, for the full-scale training data set of QuickBird, 50 PAN images with a size of  $256 \times 256$  are directly chosen from Sundarbans, India, to obtain the HR/LR filters. Similarly, the reduced-scale and full-scale training data sets of Geoeye-1 are made up of 50 simulated and real PAN images from Hobart, Australia, respectively.

### C. Experiments on Reduced-Scale Data Set

In this experiment, two pairs of simulated source images from QuickBird and Geoeye-1 are utilized to evaluate the performance of the proposed method and other methods. Fig. 4(a) and (b) presents the LR MS image with a resolution

TABLE I  
NUMERICAL EVALUATION OF FUSED RESULTS ON QUICKBIRD DATA SET

Metric	GIHS [10]	GS [7]	AWLP [12]	SVT [16]	SRLD [24]	SparseFI [26]	CSSC	
Q4	0.8781	0.8812	0.9072	0.8288	0.9047	0.9115	<b>0.9182</b>	
SAM	3.8975	3.6962	2.6095	3.9287	2.8892	<b>2.5961</b>	2.9282	
R	0.4857	0.5035	0.8891	0.7891	0.9003	0.9150	<b>0.9175</b>	
G	0.8233	0.7559	0.9223	0.9095	0.8265	0.9363	<b>0.9388</b>	
UIQI	B	0.8366	0.8900	0.9196	0.9108	0.9266	0.9349	<b>0.9389</b>
NIR	0.9573	0.9581	0.9639	0.9079	0.9616	0.9660	<b>0.9666</b>	
Avg.	0.7755	0.7769	0.9237	0.8793	0.9287	0.9381	<b>0.9405</b>	
ERGAS	3.3829	3.1959	1.9670	2.5779	2.0010	<b>1.8837</b>	2.0420	

of 11.2 m and the HR PAN image with a resolution of 2.8 m from QuickBird, respectively. Then, fused images of LR MS and HR PAN images are compared with the original HR MS image in Fig. 4(c). The fusion results of all methods are displayed in Fig. 4(d)–(j). Besides, the areas circled by red line are magnified and put in the lower right corners of fused images for a more intuitive visual comparison. From Fig. 4, obvious spectral distortions can be observed from the results of the compared methods and the proposed method when compared with the reference image. We can see that the spectral information in the results of GS [7] and GIHS [10] methods is distorted heavily because of the mismatch in the spectral range. AWLP [12] and SVT [16] perform better than GIHS and GS in spectral preservation but lose some spatial details. Besides, SRLD [24] and SparseFI [26] behave well in spectral information. However, some blurring effects can be found in the vegetation areas in Fig. 4(h)–(i), which may be caused by patch aggregation for HR MS image reconstruction. The proposed method can preserve the spectral characteristics more effectively. Then, the numerical values of the fused images in Fig. 4 are listed in Table I, where the best result for each metric is labeled in bold. R (Red), G (Green), B (Blue), and NIR (near infrared) are the bands of MS image. Avg. denotes the average values of metrics of all bands. It can be observed that the best Q4 and UIQI are provided by the

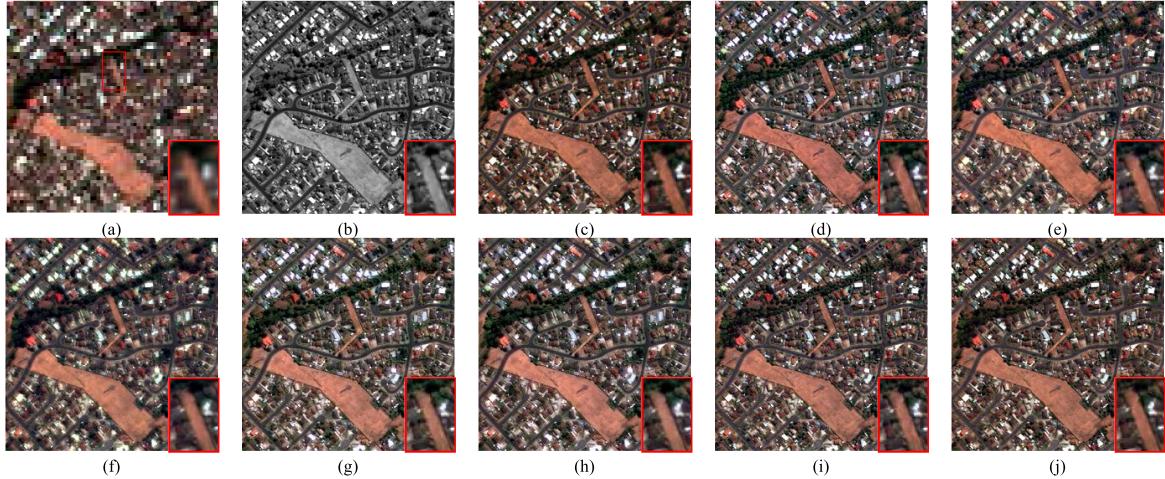


Fig. 5. Fusion results of images from Geoeye-1. (a) Resampled LR MS image. (b) HR PAN image. (c) Reference HR MS image. (d) GIHS [10]. (e) GS [7]. (f) AWLP [12]. (g) SVT [16]. (h) SRLD [24]. (i) SparseFI [26]. (j) CSSC.

TABLE II

NUMERICAL EVALUATION OF FUSED RESULTS ON GEOEYE-1 DATA SET

Metric	GIHS [10]	GS [7]	AWLP [12]	SVT [16]	SRLD [24]	SparseFI [26]	CSSC	
Q4	0.7208	0.7033	0.7633	0.7531	0.7650	0.7734	<b>0.7820</b>	
SAM	8.3160	6.3419	6.3413	6.2931	6.1256	<b>5.5914</b>	6.2618	
R	0.9102	0.8975	0.9271	0.9142	0.9538	0.9579	<b>0.9611</b>	
G	0.8610	0.9021	0.9598	0.9561	0.9685	0.9705	<b>0.9722</b>	
UIQI	B	0.8755	0.8956	0.9225	0.9426	0.9687	0.9703	<b>0.9724</b>
NIR	0.8179	0.8046	0.8319	0.8250	0.8601	0.8725	<b>0.8819</b>	
Avg.	0.8662	0.8749	0.9178	0.9095	0.9378	0.9428	<b>0.9469</b>	
ERGAS	8.1593	7.6541	3.5409	3.6825	3.3202	<b>3.1854</b>	3.3359	

proposed method, which means fewer spectral distortions. As for SAM and ERGAS, SparseFI [26] gives the best values.

Fig. 5(a) reports the LR MS image from Geoeye-1 satellite, whose spatial resolution is 8 m. The HR PAN image with a resolution of 2 m is shown in Fig. 5(b). The reference HR MS image is presented in Fig. 5(c). Fig. 5(d)–(j) shows the fused images of all methods. Some unnatural color can be found in the tree area of Fig. 5(d) and (e). In the result of SVT [16], spatial details are influenced because improper high frequencies from the HR PAN image are injected. For the sparse coding-based methods, SRLD [24] introduces some spectral distortions and the spatial details in some areas are smooth and blurring in the result of SparseFI [26]. For the proposed method, the spatial and spectral information is preserved better. Table II gives the quality assessment results of each method, and the best result is labeled in bold. Although the best values of SAM and ERGAS are from SparseFI [26], the proposed method generates the best Q4 and UIQI.

#### D. Experiments on Full-Scale Data Set

In this experiment, QuickBird and Geoeye-1 provide two pairs of real source images for subjective and objective evaluation. Fig. 6(a) and (b) shows a 2.8-m LR MS image and a 0.7-m HR PAN image from Quickbird. There is no reference HR MS image for comparison because the experiment is



Fig. 6. Fusion results of images from QuickBird. (a) Resampled LR MS image. (b) HR PAN image. (c) GIHS [10]. (d) GS [7]. (e) AWLP [12]. (f) SVT [16]. (g) SRLD [24]. (h) SparseFI [26]. (i) CSSC.

conducted at full scale. Fig. 6(c)–(i) reports the fusion results of the compared methods. Besides, specific areas in fused images are chosen and magnified to analyze the visual difference more easily. In Fig. 6, one can see spectral distortions from the tree area of Fig. 6(c) and (d), which is similar to the reduced-scale experimental results in Fig. 4. The result in Fig. 6(h) also suffers from the loss of spectral information. However, AWLP [12], SVT [16], and the proposed method presents satisfactory spectral performance. For spatial structures, it seems that the results in Fig. 6(e) and (f) lose the spatial details in the building areas. Compared with other methods, the proposed method generates better spatial quality.

TABLE III

NUMERICAL EVALUATION OF FUSED RESULTS ON QUICKBIRD DATA SET

Metric	GIHS [10]	GS [7]	AWLP [12]	SVT [16]	SRLD [24]	SparseFI [26]	CSSC
$D_\lambda$	0.0820	0.0836	0.0669	0.0659	0.0618	<b>0.0476</b>	0.0647
$D_S$	0.0909	0.0867	0.0777	0.0843	0.0814	0.0976	<b>0.0505</b>
QNR	0.8346	0.8369	0.8606	0.8562	0.8618	0.8594	<b>0.8880</b>

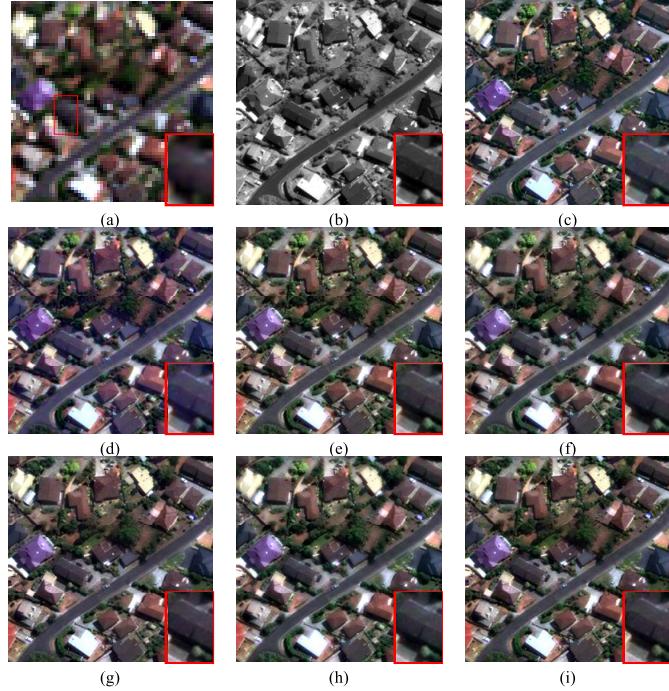


Fig. 7. Fusion results of images from Geoeye-1. (a) Resampled LR MS image. (b) HR PAN image. (c) GIHS [10]. (d) GS [7]. (e) AWLP [12]. (f) SVT [16]. (g) SRLD [24]. (h) SparseFI [26]. (i) CSSC.

Table III provides the values of  $D_S$ ,  $D_\lambda$ , and QNR for numerical comparison, in which the best result is labeled in bold. The best value of  $D_\lambda$  is generated by SparesFI [26], but our proposed method produces best  $D_S$  and QNR values.

Besides, fusion results of the real images from Geoeye-1 are shown in Fig. 7. The 2-m LR MS and 0.5-m HR PAN images are displayed in Fig. 7(a) and (b), respectively. The fused results are shown in Fig. 7(c)–(i). It can be found that some spectral distortions arise in the road area in Fig. 7(d) and the color of the area looks unnatural. AWLP [12], SVT [16], and SparseFI [26] behave well in spectral preservation, but the spatial details of tree area in Fig. 7(f) are lost. Besides, the spectral information of SRLD [24] is distorted, such as the purple building area in Fig. 7(g). Compared with other methods, the proposed method provides a better visual performance. Besides, the numerical values of all indexes are given in Table IV. The proposed method achieves the best values of  $D_\lambda$  and QNR. The best value of  $D_S$  is given by SVT [16] and followed by the proposed method.

#### E. Time Analysis

In this section, we compare the execution time of the proposed method and other compared methods. The results

TABLE IV

NUMERICAL EVALUATION OF FUSED RESULTS ON GEOEYE-1 DATA SET

Metric	GIHS [10]	GS [7]	AWLP [12]	SVT [16]	SRLD [24]	SparseFI [26]	CSSC
$D_\lambda$	0.0944	0.0868	0.1295	0.1286	0.1103	0.1086	<b>0.1054</b>
$D_S$	0.1213	0.1114	0.0785	<b>0.0567</b>	0.0602	0.0587	0.0579
QNR	0.7957	0.8115	0.8022	0.8220	0.8361	0.8390	<b>0.8428</b>

TABLE V

TIME COMPARISON OF ALL METHODS ON DIFFERENT DATA SETS (s)

Method	Reduced-Scale Dataset		Full-Scale Dataset	
	QuickBird	Geoeye-1	QuickBird	Geoeye-1
GIHS[10]	0.69	0.68	0.29	0.39
GS[7]	0.72	0.69	0.33	0.35
AWLP[12]	1.12	1.09	1.43	1.44
SVT[16]	13.91	13.22	13.30	12.25
SRLD[24]	5150.81	5168.42	5153.47	5161.91
SparseFI[26]	130.35	137.78	129.33	139.35
CSSC	1947.72	1939.20	1944.24	1944.64

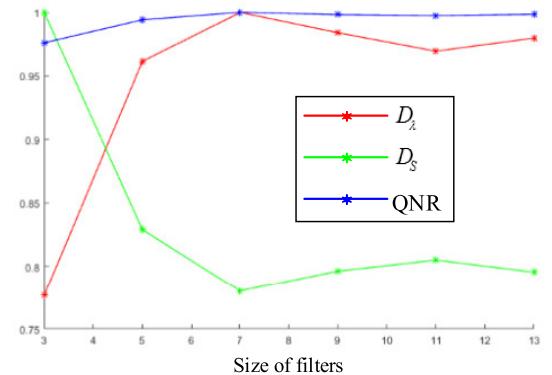


Fig. 8. Performance of the proposed method with different filter sizes on full-scale QuickBird data set.

on different data sets are reported in Table V, where the time is measured in seconds. The experiments are implemented in MATLAB R2013a on a server with Xeon CPU E5-2609/2.4 GHz/16-GB RAM. From Table V, it can be observed that the methods based on sparse representation are time consuming when compared with CS- and MRA-based methods. In SRLD [24], the dictionary is learned by K-SVD [29] first from source images and then sparse coefficients are solved to reconstruct the HR MS image. Due to the involved dictionary training, SRLD [24] takes a long time. SparseFI [26] only needs to obtain sparse coefficients of LR MS image because of the dictionary is directly constructed by PAN images. In CSSC, HR/LR filters are pretrained and the solution of linear systems takes a longer time when the source images are fused. For sparse representation-based methods, SparseFI [26] spends less time.

#### F. Effects of Size and Number of Filters

The filter size and number are the key parameters in the proposed method. First, we test the performance of the

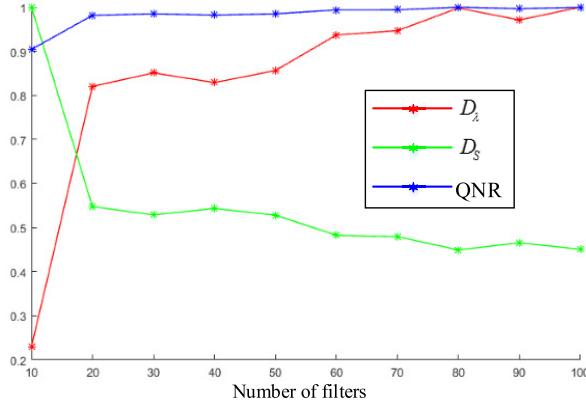


Fig. 9. Performance of the proposed method with different numbers of filters on full-scale QuickBird data set.

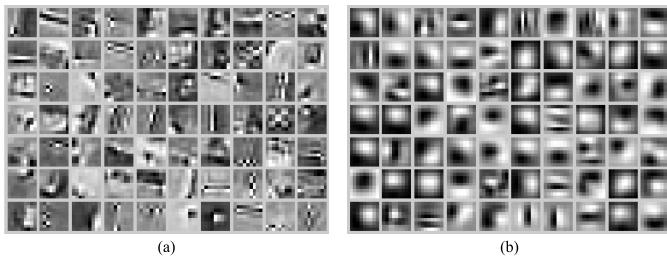


Fig. 10. Visualization of HR/LR filters learned by the proposed learning framework. (a) HR filters. (b) LR filters.

proposed method for different filter sizes. Fig. 8 shows the curve of each index on full-scale QuickBird data set, and the values are normalized for direct comparison. In Fig. 8, the horizontal axis is the size of filter and the size of filter varies from 3 to 13 with step 2. We know that the filters with a large size can contain useful information but ignore the intrinsic structures. Conversely, filters with small size have a limited capability to learn the features. It also can be observed that poor values of  $D_S$  and QNR are obtained when the filter size is 3 in Fig. 8. With the increasing of filter size, QNR becomes better, but the difference is small with different filter sizes. Therefore, the size of filter is set as  $9 \times 9$  considering the overall performance.

Then, we also analyze the influence of different numbers of filters on fusion results. Fig. 9 shows the variations of all indexes for full-scale QuickBird data set with the increasing number of filters, where the horizontal axis is the number of filters with the step 10. In Fig. 9, we can find that the proposed method behaves poorly with fewer filters on QNR and  $D_S$ . Because it is difficult to adequately characterize the spatial structures in the image with a small number of filters. For a larger filter number, the sparsity in feature maps can be better captured, but the proposed method suffers from higher time and space complexity. Moreover,  $D_\lambda$  increases with the increasing number of the filter. Thus, the number of filters is set as 70 considering the tradeoff between performance and complexity.

Besides, learned HR/LR filters from full-scale QuickBird data set are also shown in Fig. 10 for a more intuitive analysis.

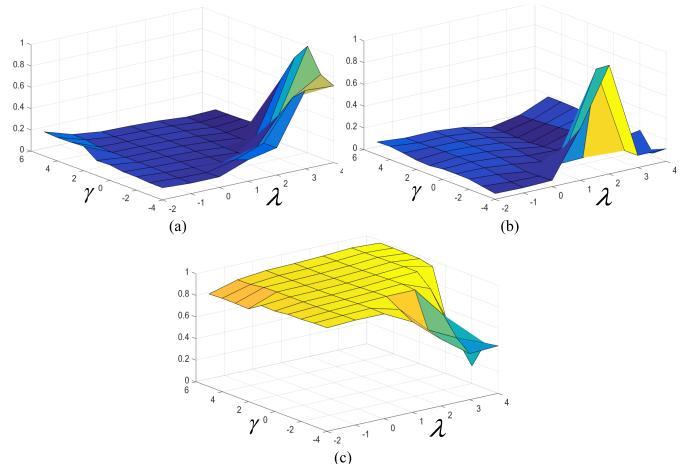


Fig. 11. Performance of the proposed method with the variations of  $\gamma$  and  $\lambda$  on full-scale QuickBird data set. (a)  $D_\lambda$ . (b)  $D_S$ . (c) QNR.

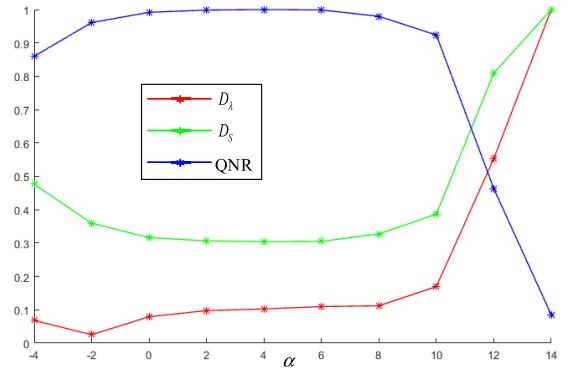


Fig. 12. Performance of the proposed method with the variation of  $\alpha$  on full-scale QuickBird data set.

HR filters in Fig. 10(a) are distributed over orientation and contain some center surround and edges, which can efficiently enhance the spatial details of the HR MS images. LR filters in Fig. 10(b) are blurry compared with the HR filters because they are inferred from the LR PAN images. A good correspondence to the HR filters can be found in LR filters, which helps to obtain more reasonable feature maps.

#### G. Effects of Regularization Parameters

In the fusion model, there are three regularization parameters  $\alpha$ ,  $\lambda$ , and  $\gamma$ .  $\alpha$  controls the reconstruction fidelity of the HR PAN image. For the smooth term, larger  $\gamma$  leads to the appearance of much low frequency information in feature maps.  $\lambda$  controls the sparsity term, which has an important influence on the fusion results. To select appropriate values for the regularization parameters, the performance of the proposed method on full-scale QuickBird data set is evaluated with the variations of these parameters.

Fig. 11 shows the normalized results of  $D_\lambda$ ,  $D_S$  and QNR for different pairs  $\gamma$  and  $\lambda$ . In Fig. 11,  $\gamma$  and  $\lambda$  vary from  $4^{-4}$  to  $4^5$  and from  $4^{-2}$  to  $4^4$  with step 1 in power, respectively. The  $x$ -axis and  $y$ -axis in Fig. 11 record the log base 4 of  $\gamma$  and  $\lambda$ . It can be seen that the numerical results vary with the variations of  $\gamma$  and  $\lambda$  in Fig. 11. QNR increases first and

then decreases with the increasing of  $\lambda$ , and the best value is achieved when  $\lambda = 4^3$ . The proposed method produces better  $D_\lambda$  when  $\lambda$  is  $4^0$ . Considering the overall performance of the proposed method,  $\gamma$  and  $\lambda$  are set as 4 and 4, respectively.

The normalized results of  $D_\lambda$ ,  $D_S$ , and QNR are displayed in Fig. 12.  $\alpha$  varies from  $2^{-4}$  to  $2^{14}$  with the step 2 in power, and the abscissa values are the log base 2 of  $\alpha$ . We can see the best QNR is achieved when  $\alpha = 2^4$ .  $D_\lambda$  increases with the increasing of  $\alpha$ . Considering the balance with the fidelity term of LR MS image, we set  $\alpha$  as  $2^3$ . According to the same experiments,  $\gamma$ ,  $\lambda$ , and  $\alpha$  are set as 1, 4, and  $2^4$  for Geoeye-1 data sets.

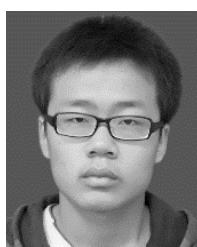
## V. CONCLUSION

In this paper, we proposed a novel fusion method for LR MS and HR PAN images based on CSSC. The CSC model is combined with the degradation of source images to avoid the patch partition. In the model, feature maps of the HR MS image over HR/LR filters are calculated by using LR MS and HR PAN images simultaneously. Moreover, the strong correlation among bands is depicted by structure sparsity to ensure the spectral structure in the HR MS image. Besides, we also establish a new joint HR/LR filter learning framework. Through BCD algorithm, the HR/LR filters are inferred from HR/LR PAN image pairs by sharing feature maps in this scheme. Compared with some classical methods on different data sets, the proposed fusion method can obtain better results with less spatial distortions. For future work, we will find more reasonable spectral prior, such as local spectral similarity, to preserve the spectral information in the fused image.

## REFERENCES

- [1] W. Zhao *et al.*, “Superpixel-based multiple local CNN for panchromatic and multispectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 4141–4156, Jul. 2017.
- [2] D. Wen, X. Huang, L. Zhang, and J. A. Benediktsson, “A novel automatic change detection method for urban high-resolution remotely sensed imagery based on multiindex scene representation,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 609–625, Jan. 2016.
- [3] J. Han, D. Zhang, G. Cheng, L. Guo, and J. Ren, “Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3325–3337, Jun. 2015.
- [4] G. Vivone *et al.*, “A critical comparison among pansharpening algorithms,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, May 2015.
- [5] C. Thomas, T. Ranchin, L. Wald, and J. Chanussot, “Synthesis of multispectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics,” *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1301–1312, May 2008.
- [6] Z. Wang, D. Ziou, C. Armenakis, D. Li, and Q. Li, “A comparative analysis of image fusion methods,” *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1391–1402, Jun. 2005.
- [7] 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.
- [8] 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.
- [9] W. Carper, T. Lillesand, and R. Kiefer, “The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multispectral image data,” *Photogramm. Eng. Remote Sens.*, vol. 56, no. 4, pp. 459–467, Apr. 1990.
- [10] 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.
- [11] T. Ranchin and L. Wald, “Fusion of high spatial and spectral resolution images: The ARSIS concept and its implementation,” *Photogramm. Eng. Remote Sens.*, vol. 66, no. 1, pp. 49–61, Jan. 2000.
- [12] 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.
- [13] J. Nunez, X. Otazu, O. Fors, A. Prades, V. Pala, and R. Arbiol, “Multiresolution-based image fusion with additive wavelet decomposition,” *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1204–1211, May 1999.
- [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. 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.
- [16] S. Zheng, W.-Z. Shi, J. Liu, and J. Tian, “Remote sensing image fusion using multiscale mapped LS-SVM,” *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1313–1322, May 2008.
- [17] M. Ghahremani and H. Ghassemian, “A compressed-sensing-based pan-sharpening method for spectral distortion reduction,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 2194–2206, Apr. 2016.
- [18] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, “A new pansharpening algorithm based on total variation,” *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 318–322, Jan. 2014.
- [19] D. Fasbender, J. Radoux, and P. Bogaert, “Bayesian data fusion for adaptable image pansharpening,” *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 6, pp. 1847–1857, Jun. 2008.
- [20] 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.
- [21] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [22] S. Li and B. Yang, “A new pan-sharpening method using a compressed sensing technique,” *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 738–746, Feb. 2011.
- [23] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.
- [24] 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.
- [25] C. Jiang, H. Zhang, H. Shen, and L. Zhang, “A practical compressed sensing-based pan-sharpening method,” *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 4, pp. 629–633, Jul. 2012.
- [26] 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.
- [27] X. X. Zhu, C. Grohnfeldt, and R. Bamler, “Exploiting joint sparsity for pansharpening: The J-SparseFI algorithm,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 2664–2681, May 2016.
- [28] C. Jiang, H. Zhang, H. Shen, and L. Zhang, “Two-step sparse coding for the pan-sharpening of remote sensing images,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 5, pp. 1792–1805, May 2014.
- [29] M. Aharon, M. Elad, and A. Bruckstein, “K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation,” *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
- [30] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, “Deconvolutional networks,” in *Proc. IEEE CVPR*, San Francisco, CA, USA, Aug. 2010, pp. 2528–2535.
- [31] H. Bristow, A. Eriksson, and S. Lucey, “Fast convolutional sparse coding,” in *Proc. IEEE CVPR*, Portland, OR, USA, Jun. 2013, pp. 391–398.
- [32] B. Wohlberg, “Efficient algorithms for convolutional sparse representations,” *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 301–315, Jan. 2016.
- [33] S. Gu, W. Zuo, Q. Xie, D. Meng, X. Feng, and L. Zhang, “Convolutional sparse coding for image super-resolution,” in *Proc. IEEE ICCV*, Santiago, Chile, Dec. 2016, pp. 1823–1831.

- [34] F. Heide, W. Heidrich, and G. Wetzstein, "Fast and flexible convolutional sparse coding," in *Proc. IEEE CVPR*, Boston, MA, USA, Jun. 2015, pp. 5135–5143.
- [35] L. Yang *et al.*, "Image reconstruction via manifold constrained convolutional sparse coding for image sets," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 7, pp. 1072–1081, Oct. 2017.
- [36] 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.
- [37] K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. Lecun, "What is the best multi-stage architecture for object recognition?" in *Proc. IEEE ICCV*, Kyoto, Japan, Sep./Oct. 2010, pp. 2146–2153.
- [38] 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.
- [39] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM J. Imag. Sci.*, vol. 6, no. 3, pp. 1758–1789, 2013.
- [40] A. Kallel, "MTF-adjusted pansharpening approach based on coupled multiresolution decompositions," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3124–3145, Jun. 2015.
- [41] 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.
- [42] 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.
- [43] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.
- [44] 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.
- [45] 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.
- [46] S. Hassani, *Mathematical Methods: For Students of Physics and Related Fields*. New York, NY, USA: Springer, 2000, pp. 289–319.
- [47] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang, "Coupled dictionary training for image super-resolution," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3467–3478, Aug. 2012.
- [48] A. Krause and V. Cevher, "Submodular dictionary selection for sparse representation," in *Proc. ICML*, Haifa, Israel, Jun. 2010, pp. 567–574.
- [49] 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.
- [50] Q. Pan, Y. Liang, L. Zhang, and S. Wang, "Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis," in *Proc. IEEE CVPR*, Providence, RI, USA, Jun. 2012, pp. 2216–2223.
- [51] S. Xiang, G. Meng, Y. Wang, C. Pan, and C. Zhang, "Image deblurring with coupled dictionary learning," *Int. J. Comput. Vis.*, vol. 114, nos. 2–3, pp. 248–271, 2015.



**Kai Zhang** received the B.S. degree in electrical engineering and automation from the North University of China, Taiyuan, China, in 2013. He is currently pursuing the Ph.D. degree in circuits and systems with the Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education, Xidian University, Xi'an, China.

His research interests include machine learning, sparse representation, and image fusion.



**Min Wang** received the B.S. degree in autocontrol and the M.S. and Ph.D. degrees in signal and information processing from Xidian University, Xi'an, China, in 2000, 2003, and 2005, respectively.

He is currently an Associate Professor with the National Laboratory of Radar Signal Processing, Xidian University. His research interests include radar signal processing, statistical signal processing, and impulse radio.



**Shuyuan Yang** (SM'14) received the B.S. degree in electrical engineering and the M.S. and Ph.D. degrees in circuit and system from Xidian University, Xi'an, China, in 2000, 2003, and 2005, respectively.

Her research interests include intelligent signal processing, machine learning, and image processing.



**Licheng Jiao** (SM'89–F'17) received the B.S. degree from Shanghai Jiao Tong University, Shanghai, China, in 1982, and the M.S. and Ph.D. degrees from Xi'an Jiaotong University, Xi'an, China, in 1984 and 1990, respectively.

He is currently a Distinguished Professor with the Key Laboratory of Intelligent Perception and Image Understanding, Ministry of Education of China, Xidian University, Xi'an, where he is also a Professor and the Dean of the Electronic Engineering School.

He has led approximately 40 important scientific research projects. He has authored over 10 monographs, 100 papers in international journals and conferences, and three books: *Theory of Neural Network Systems* (Xidian University Press, 1990), *Theory and Application on Nonlinear Transformation Functions* (Xidian University Press, 1992), and *Applications and Implementations of Neural Networks* (Xidian University Press, 1996). He has authored or co-authored over 150 scientific papers. His research interests include signal and image processing, natural computation, intelligent information processing neural networks, data mining, nonlinear intelligence signal processing, and communication.