Fusion of Panchromatic and Multispectral Images Using Multiscale Convolution Sparse Decomposition

Kai Zhang, Feng Zhang, Zhixi Feng, Jiande Sun, and Quanyuan Wu

***Abstract*—In this article, we proposed a novel image fusion method based on multiscale convolution sparse decomposition (MCSD). A unified framework based on MCSD is first utilized to decompose panchromatic (PAN) image and the spatial component of upsampled low spatial resolution multispectral (LR MS) images, which can produce the corresponding low frequencies and feature maps. By combining convolution sparse decomposition with multiscale analysis, MCSD can efficiently approximate the spatial and spectral information in images. Next, a binary map generated from gradient information is utilized to integrate the low frequencies of LR MS and PAN images. For feature maps, the fusion gain for each pixel is calculated according to the similarity between the local patches from them. Finally, the fused image is reconstructed by the sum of fused low frequency and feature maps. Some experiments are conducted on QuickBird and GeoEye-1 satellite datasets. Compared with other methods, the proposed method performs better in visual and numerical evaluations.**

***Index Terms*—Convolution sparse representation (SR), image fusion, multiscale decomposition, multispectral (MS) image, panchromatic (PAN) image.**

I. Introduction

In RECENT years, with the rapid development of imaging techniques, a series of remote sensing satellites have been launched to achieve more comprehensive earth observation. So more and more remote sensing images have been widely used on many fields for scene interpretation, such as object detection [1], classification [2], and change detection [3]. However, there is a fundamental tradeoff between spatial and spectral resolution for remote sensing images due to the limitation of signal-noise ratio. For instance, for multispectral (MS) image, its spatial resolution is lower than that of panchromatic (PAN) image, but MS image contains abundant spectral information when compared with PAN image composed of only one channel. Hyperspectral image provides more rich spectral signatures but spatial resolution is also low. Therefore, image fusion techniques [4]–[7] are proposed to improve the spatial and spectral resolution of remote sensing images.

Over the past two decades, a variety of methods are proposed for the fusion of PAN and MS images and achieved satisfactory fusion results, which can be separated as four types: component substitution based methods, multiresolution analysis (MRA) [8] based methods, spatial–spectral degradation model-based methods and deep neural network (DNN) based methods. For the first kind of methods, the observed low spatial resolution MS (LR MS) image is interpolated to match the size of PAN image. Then, some transforms are adopted to estimate the component which contains most of the spatial information in LR MS image. PAN image is used to substitute the spatial component. Finally, the corresponding inverse transform is implemented on the synthesized spatial component and other components to obtain the fused HR MS image. According to the framework, Intensity-Hue-Saturation transform [9], principal component analysis (PCA) [9] and Gram–Schmidt (GS) transform [10] are considered to fuse PAN and LR MS images. These methods are widely used because of the simple principle and high efficiency. However, spectral distortion can be always found in the fusion results due to the spectral response range differences between LR MS and PAN images. Thus, some methods [11]–[13] are proposed to alleviate the issue. For example, Kim *et al*. [11] employed spatial PCA to obtain more reasonable spatial structures from the bands of LR MS image.

For the second kind of methods, they assume that the missing spatial details in LR MS image can be found in PAN image, which is denoted as *Amélioration de la Résolution Spatiale par Injection de Structures* (ARSIS) [14]. Then, MRA methods are utilized to extract the spatial details from PAN image, which are injected into the interpolated LR MS image. For example, Otazu *et al*. [15] took the physical electromagnetic spectrum responses into consideration and estimated more reasonable spatial details using wavelet transform. Alparone *et al*. [16] demonstrated that MRA-based pansharpening is described by a separable low-pass filter derived from modulation transfer function (MTF) of MS images. Vivone *et al*. [17]–[19] exploited deeply different regression models to calculate more accurate injection coefficients for the spatial information enhancement of MS images. Besides, Zheng *et al*. [20] developed the support

Manuscript received June 11, 2020; revised July 26, 2020, August 27, 2020, and November 16, 2020; accepted December 4, 2020. Date of publication December 9, 2020; date of current version January 6, 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 Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, Xidian University under Grant IPIU2019008, in part by the Natural Science Foundation for Distinguished Young Scholars of Shandong Province under Grant JQ201718, and in part by the State Key Program of National Natural Science of China under Grant 61836009. *(Corresponding author: Kai Zhang.)*

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

Zhixi Feng is with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, Xidian University, Xi'an 710071, China (e-mail: zhxfeng2013@gmail.com).

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.2020.3043521

value transform (SVT) to realize the fusion of PAN and LR MS images. Subsequently, Yang *et al*. [21] further obtained better fusion results by combining SVT with adaptive PCA. These methods can better preserve the spectral information in HR MS image. But some spatial artifacts are caused by the excessive injection of spatial details from PAN image.

For the third kind of fusion methods, it is supposed that the acquired PAN and LR MS images are the degraded results of HR MS images in spatial and spectral domains, respectively. Then, the fusion of PAN and LR MS images is conducted by the degradation model with reasonable regularized priors. For example, Li *et al*. [22] considered the sparse prior and proposed a new pansharpening technique based on compressive sensing theory [23]. Besides, some improved versions [24]–[25] are presented to generate better fusion results and make the model more practical. For example, Wang *et al*. [24] introduced local autoregressive model into the model to improve the spatial structures of HR MS image. Besides, other efficient priors [26]–[31], such as total-variation [26], non-negativity [27], and low-rank property [28]–[29], are further considered. For example, Yang *et al*. [28] formulated LR MS image as the sum of HR MS image and two different images, in which low-rank prior is utilized to capture the spatial and spectral similarity in HR MS image. Besides, sparsity-induced priors are also considered in [32]–[34] for pansharpening. These methods behave well in spectral preservation and spatial enhancement. However, high computational complexity cannot be ignored.

Besides, motivated by the great achievements in remote sensing target detection and classification [35]–[37] of DNN, many image fusion methods based on DNN gradually appear. For example, Huang *et al*. [38] first considered modified sparse denoising autoencoder to model the relationship between HR and LR images for fusion of LR MS and PAN images. Then, Masi *et al*. [39] adopted convolution neural network (CNN) to fuse LR MS and PAN images. Subsequently, a target-adaptive CNN [40] was designed to further improve the reconstruction accuracy of fused image. Moreover, Shao *et al*. [41] also used CNN and designed two CNNs with different architectures to extract the spatial and spectral information in PAN and LR MS images. In [42], high frequencies are inferred from DNN and then injected into the upsampled LR MS image, which follows the concept of ARSIS. Wei *et al*. [43]–[44] proposed a deep residual pansharpening neural network (DRPNN) to boost the accuracy of the fusion results. Besides, Wang *et al*. [45] introduced dense blocks into CNN and residual learning is considered for spatial resolution enhancement.

Recently, convolution sparse decomposition (CSD) [46]–[47] is gaining much attention for fusion of PAN and LR MS images because of global representation and better reconstruction performance compared with sparse representation (SR). For example, Zhang *et al*. [48] combined CSD with spatial–spectral degradation model and introduced structural sparse prior to capture the spectral correlation in the bands of MS image. In [49], CSD was employed to decompose the high frequencies of source images and the corresponding feature maps are fused. Fei *et al*. [50] adopted the filter sharing strategy in different dictionaries to synthesize the spatial details which are injected into

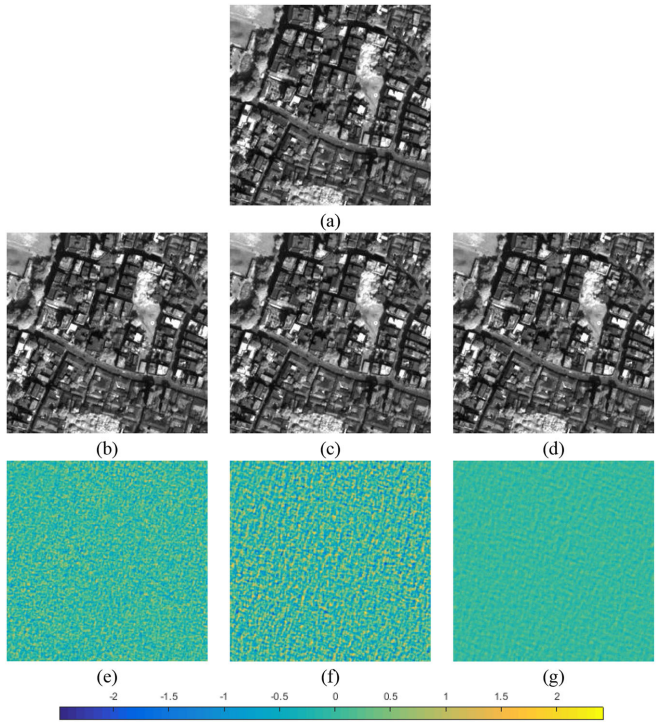

Fig. 1. Reconstruction performance with different scales. (a) Original PAN image. (b) Reconstructed PAN image by 10 filters with size 5×5. (c) Reconstructed PAN image by 10 filters with size 9×9. (d) Reconstructed PAN image by five filters with size 5×5 and five filters with size 9×9. (e)–(g) Error images of (b), (c), (d), respectively.

fused images. Although some good fusion results are achieved, these methods only use filters with the same scale to analyze the spatial details in PAN and MS images, which ignore the multiscale property in these images. Thus, some subtle or finer spatial features cannot be decomposed well by the filters with the same scale [51]. Fig. 1 shows the reconstruction performance of single-scale and multiscale filters for the given PAN image with size 256×256 from QuickBird satellite. Although, we can see that the PAN image can be reconstructed well by single-scale and multiscale filters in Fig. 1(b)–(c), it can be found that multiscale filters can more accurately approximate to the original image in spatial details from the error images in Fig. 1(e)–(g). Thus, compared with the single-scale CSD, multiscale convolution sparse decomposition (MCSD) should be further explored to more accurately represent the images. Besides, due to statistical characteristic differences with different scales, the same fusion rule for feature maps from different scales will result in some spatial and spectral distortions. So specific fusion rule should be designed for different scales.

Taking the above two issues into consideration, an image fusion method based on MCSD is developed by combining the elaborate feature map fusion rules for different scales. In the fusion method, GS transform is carried out on LR MS image to obtain the spatial component. Then, we establish a unified framework to simultaneously achieve the high and low frequency separation and MCSD of high frequency for PAN image or the spatial component from LR MS image. CSD is extended to MCSD to approximate the images to be fused for

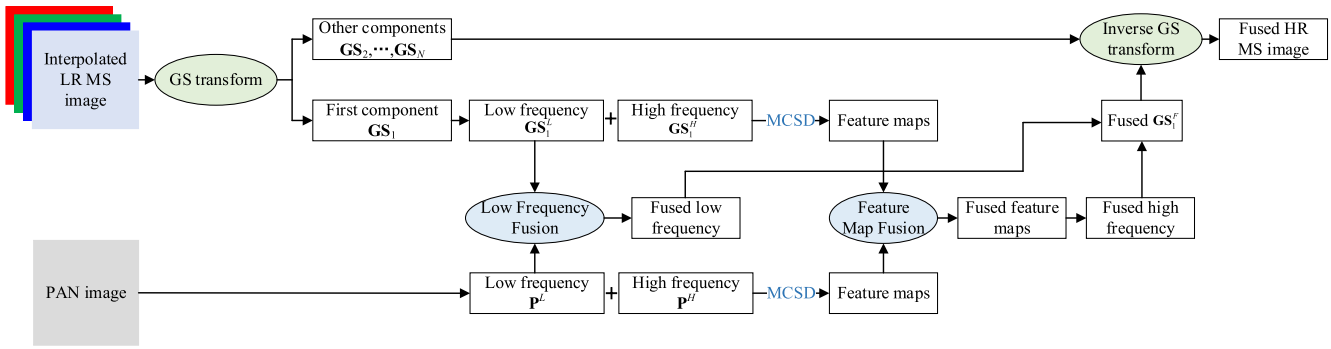

Fig. 2. Flowchart of the proposed method based on MCSD.

effective spatial information representation, which is solved by alternating direction method of multipliers (ADMM) [52]. Next, different fusion rules are designed for low frequencies and the feature maps from high frequencies. For low frequencies of PAN image and the spatial component of LR MS image, a binary map is inferred from the gradient and intensity information to guide the fusion of low frequencies, which can further enhance the spatial details in the fused image. For the feature maps on different scales, local similarity is calculated as a combination weight for each pixel by considering the scale differences. Finally, the fused image can be produced by combining low and high frequencies after reconstruction of the fused feature maps. Experiments on full-resolution and reduced-resolution datasets demonstrate that the proposed method provides better visual and numerical performance. Some parameters about MCSD are also investigated to analyze the influences on the fusion results. The contributions of this article are listed as the following two points.

1) By combining multiscale analysis with CSD, MCSD is developed and then a unified framework is established to separate the low and high frequencies of LR MS and PAN images, where MCSD is employed to decompose the high frequencies more accurately.
2) Considering the spatial information differences among different feature maps from different scales, the injection gain for each pixel is computed from the similarity between the local patches from LR MS and PAN images, which avoids the spectral distortions caused by global gain.

The remainder of the article is organized as follows. The image fusion framework based on MCSD is introduced in Section II. Section III describes MCSD and the optimization algorithm in detail. In Section IV, experimental results and comparisons on full-resolution and reduced-resolution datasets are presented. Conclusions are described in Section V.

II. PAN and LR MS Image Fusion Based on MCSD

In this section, we apply MCSD on PAN and LR MS images and integrate them to obtain HR MS image. In the proposed method, GS transform is first implemented on the LR MS image to obtain the first component $\mathbf{GS}_1 \in \mathcal{R}^{h\times w}$. h and w are the height and width of images. For $\mathbf{GS}_1$ and PAN image, a unified approximation is next designed to achieve the high- and low-frequency separation and fine-grained feature map estimation of high frequencies. Then, the corresponding feature maps of high frequencies are merged by different fusion rules derived from multiscale property. Low frequencies of $\mathbf{GS}_1$ and PAN image are fused by combining their gradient information. Subsequently, the fused spatial component $\mathbf{GS}_1^F \in \mathcal{R}^{h\times w}$ is synthesized by combining the reconstructed high frequency from the fused feature maps and the fused low frequency. Finally, the HR MS image is generated through inverse GS transform. The flowchart of the proposed method based on MCSD is shown in Fig. 2 and the detailed procedures are introduced in the following parts.

A. High-/Low-Frequency Separation and Feature Map Decomposition

Generally, the spatial details also termed as high frequencies from PAN image are injected into the obtained first component $\mathbf{GS}_1$ after GS transform. The high- and low-frequency separation of the image has an effect on the spectral distortions in the final fused results because the spectral information is generally influenced by low-frequency component. Besides, high frequencies contain some obvious edges, curves, and textures. In order to estimate the smoothing component, the L_2 norm on gradient of the low frequency is constrained. For the high frequency, direct fusion on high frequency $\mathbf{P}^H \in \mathcal{R}^{h\times w}$ of PAN image and high frequency $\mathbf{GS}_1^H \in \mathcal{R}^{h\times w}$ of the first GS component also can enhance the spatial information in LR MS image, which will lead to some spectral distortions in the fused image. Compared with the fusion on original coarse scale, decomposed feature maps on finer scales can provide a more accurate representation. Thus, MCSD is implemented on $\mathbf{P}^H$ and $\mathbf{GS}_1^H$ to compute the feature maps on different filter size.

Then, by integrating MCSD, a unified framework is established to simultaneously achieve the separation of high/low frequencies and multiscale feature map decomposition. With PAN image $\mathbf{P} \in \mathcal{R}^{h\times w}$ as the example, the unified framework is formulated as

$$\arg\min_{\mathbf{P}^L,\mathbf{Z}_{l,k}^P} \frac{1}{2}\left\|\mathbf{P}-\mathbf{P}^L-\sum_{l=1}^{L}\sum_{k=1}^{K_l}\mathbf{F}_{l,k}*\mathbf{Z}_{l,k}^P\right\|_F^2 + \frac{\alpha}{2}\left\|\nabla\mathbf{P}^L\right\|_2+\beta\sum_{l=1}^{L}\sum_{k=1}^{K_l}\left\|\mathbf{Z}_{l,k}^P\right\|_1 \quad (1)$$

where the high-frequency $\mathbf{P}^H$ of PAN image equals to $\sum_{l=1}^{L}\sum_{k=1}^{K_l}\mathbf{F}_{l,k}*\mathbf{Z}_{l,k}^P$. L and K_l stand for the number of scales and the number of filters in the lth scale, respectively. So $\mathbf{Z}_{l,k}^P \in \mathcal{R}^{h\times w}$ denotes the kth feature map in the lth scale and $\mathbf{F}_{l,k}\in\mathcal{R}^{s_l\times s_l}$ is the corresponding filter. $s_l\times s_l$ is the size of $\mathbf{F}_{l,k}$ in the lth scale. For different scales, the filter size is different. More clearly, if a multiscale filter bank is composed of three filters with size 7×7, five filters with size 9×9, and eight filters with size 11×11, the bank contains 3 scales and 16 filters. $\mathbf{P}^L\in\mathcal{R}^{h\times w}$ denotes the low frequency of PAN image. ∇ stands for the gradient operator. $*$ is the convolution operation. α and β are the tradeoff parameters. It is straightforward to adopt alternate iterative optimization algorithm to solve (1) by ADMM. After introducing an auxiliary variable $\mathbf{X}_{l,k}^P\in\mathcal{R}^{h\times w}$ for $\mathbf{Z}_{l,k}^P$, the augmented Lagrange function is written as

$$\begin{aligned}\underset{\mathbf{P}^L,\mathbf{Z}_{l,k}^P,\mathbf{X}_{l,k}^P}{\arg\min}\ &\tfrac{1}{2}\left\|\mathbf{P}-\mathbf{P}^L-\sum_{l=1}^{L}\sum_{k=1}^{K_l}\mathbf{F}_{l,k}\right.\\ &\left.*\mathbf{X}_{l,k}^P\right\|_F^2+\tfrac{\alpha}{2}\left\|\nabla\mathbf{P}^L\right\|_2+\beta\sum_{l=1}^{L}\sum_{k=1}^{K_l}\left\|\mathbf{Z}_{l,k}^P\right\|_1\\ &+\tfrac{\mu}{2}\sum_{l=1}^{L}\sum_{k=1}^{K_l}\left\|\mathbf{X}_{l,k}^P-\mathbf{Z}_{l,k}^P+\tfrac{\mathbf{Y}_{l,k}^P}{\mu}\right\|_2^2\end{aligned}\tag{2}$$

where $\mathbf{Y}_{l,k}^P\in\mathcal{R}^{h\times w}$ is a Lagrange multiplier. Then, $\mathbf{Z}_{l,k}^P$, $\mathbf{X}_{l,k}^P$, and $\mathbf{P}^L$ are updated in sequence.

For $\mathbf{Z}_{l,k}^P$, the subfunction is

$$\begin{aligned}&F_{\mathbf{Z}_{l,k}^P}\\ &=\beta\sum_{l=1}^{L}\sum_{k=1}^{K_l}\left\|\mathbf{Z}_{l,k}^P\right\|_1+\frac{\mu}{2}\sum_{l=1}^{L}\sum_{k=1}^{K_l}\left\|\mathbf{Z}_{l,k}^P-\left(\mathbf{X}_{l,k}^P+\frac{\mathbf{Y}_{l,k}^P}{\mu}\right)\right\|_2^2.\end{aligned}\tag{3}$$

Then, $\mathbf{Z}_{l,k}^P$ is updated by the soft-thresholding shrinkage operator [55] as shown below.

$$\mathbf{Z}_{l,k}^P=\mathcal{S}_{\frac{\beta}{\mu}}\left(\mathbf{R}_{l,k}^P\right)\tag{4}$$

where $\mathbf{R}_{l,k}^P=\mathbf{X}_{l,k}^P+\mathbf{Y}_{l,k}^P/\mu$, r is the element in $\mathbf{R}_{l,k}^P$ and $\mathcal{S}_{\frac{\beta}{\mu}}(r)=\operatorname{sign}(r)\cdot\max(0,|r|-\frac{\beta}{\mu})$.

For $\mathbf{X}_{l,k}^P$, its subfunction is

$$\begin{aligned}&F_{\mathbf{X}_{l,k}^P}\\ &=\frac{1}{2}\left\|\mathbf{P}-\mathbf{P}^L-\sum_{l=1}^{L}\sum_{k=1}^{K_l}\mathbf{F}_{l,k}*\mathbf{X}_{l,k}^P\right\|_F^2\\ &+\frac{\mu}{2}\sum_{l=1}^{L}\sum_{k=1}^{K_l}\left\|\mathbf{X}_{l,k}^P-\mathbf{Z}_{l,k}^P+\frac{\mathbf{Y}_{l,k}^P}{\mu}\right\|_2^2.\end{aligned}\tag{5}$$

Equation (5) is optimized after FT owing to convolution operation and the subfunction about $\mathbf{X}_{l,k}^P$ in frequency domain can be defined as

$$F_{\hat{\mathbf{x}}}=\frac{1}{2}\left\|\hat{\mathbf{p}}-\sum_{l=1}^{L}\sum_{k=1}^{K_l}\mathbf{H}_{l,k}\hat{\mathbf{x}}_{l,k}^P\right\|_2^2+\frac{\mu}{2}\sum_{l=1}^{L}\sum_{k=1}^{K_l}\left\|\hat{\mathbf{x}}_{l,k}^P-\hat{\mathbf{t}}_{l,k}^P\right\|_2^2\tag{6}$$

where $\hat{\mathbf{p}}\in\mathcal{R}^{hw}$, $\hat{\mathbf{x}}_{l,k}^P\in\mathcal{R}^{hw}$, and $\hat{\mathbf{t}}_{l,k}^P\in\mathcal{R}^{hw}$ are the vectorization versions of $\mathbf{P}-\mathbf{P}^L$, $\mathbf{X}_{l,k}^P$, and $\mathbf{T}_{l,k}^P$ after Fourier Transform (FT), respectively. $\mathbf{T}_{l,k}^P$ equals to $\mathbf{Z}_{l,k}^P-\mathbf{Y}_{l,k}^P/\mu$. $\mathbf{H}_{l,k}\in\mathcal{R}^{hw\times hw}$ is a diagonal matrix, whose diagonal elements are made up of the result of $\mathbf{F}_{l,k}$ after FT. By rearranging all feature maps into one vector, (6) is reformulated as

$$F_{\hat{\mathbf{x}}}=\frac{1}{2}\left\|\hat{\mathbf{p}}-\mathbf{H}\hat{\mathbf{x}}^P\right\|_2^2+\frac{\mu}{2}\left\|\hat{\mathbf{x}}^P-\hat{\mathbf{t}}^P\right\|_2^2\tag{7}$$

where $\hat{\mathbf{t}}^P\in\mathcal{R}^{hw(K_1+\cdots+K_L)}$ and $\hat{\mathbf{x}}^P\in\mathcal{R}^{hw(K_1+\cdots+K_L)}$ are the cascaded results of all $\hat{\mathbf{t}}_{l,k}^P$ and all $\hat{\mathbf{x}}_{l,k}^P$, respectively. $\mathbf{H}=[\mathbf{H}_{1,1},\ldots,\mathbf{H}_{l,k},\ldots,\mathbf{H}_{L,K_L}]\in\mathcal{R}^{hw\times hw(K_1+\cdots+K_L)}$. It is obvious that there is a closed-form solution for $\hat{\mathbf{x}}^P$. The derivative of (7) with respect to $\hat{\mathbf{x}}^P$ is

$$\frac{\partial F_{\hat{\mathbf{x}}^P}}{\partial\hat{\mathbf{x}}^P}=\mathbf{H}^H\mathbf{H}\hat{\mathbf{x}}^P-\mathbf{H}^H\hat{\mathbf{p}}+\mu\hat{\mathbf{x}}^P-\mu\hat{\mathbf{t}}^P\tag{8}$$

where the complex conjugate transpose is denoted by H. So the optimal value of $\hat{\mathbf{x}}^P$ is obtained by setting (8) to zero, which can be efficiently computed by Sherman–Morrison operation proposed in [47].

For $\mathbf{P}^L$, the subfunction is

$$F_{\mathbf{P}^L}=\frac{1}{2}\left\|\mathbf{P}-\mathbf{P}^L-\sum_{l=1}^{L}\sum_{k=1}^{K_l}\mathbf{F}_{l,k}*\mathbf{X}_{l,k}^P\right\|_F^2+\frac{\alpha}{2}\left\|\nabla\mathbf{P}^L\right\|_2.\tag{9}$$

So a closed-form solution for the quadratic problem can be obtained after FT

$$\mathbf{P}^L=\mathcal{F}^{-1}\left(\frac{\hat{\mathbf{G}}}{\alpha\left(\mathcal{F}(\nabla_h)^H\mathcal{F}(\nabla_h)+\mathcal{F}(\nabla_v)^H\mathcal{F}(\nabla_v)\right)+1}\right)\tag{10}$$

where $\hat{\mathbf{G}}=\mathcal{F}(\mathbf{P}-\sum_{l=1}^{L}\sum_{k=1}^{K_l}\mathbf{F}_{l,k}*\mathbf{X}_{l,k}^P)\in\mathcal{R}^{h\times w}$. $\mathcal{F}$ denotes FT and $\mathcal{F}^{-1}$ is the corresponding inverse transform of FT. ∇_h and ∇_v stand for the gradient operators in horizontal and vertical directions. Finally, $\mathbf{Y}_{l,k}^P$ is updated. When the maximum iteration number or minimum reconstruction error is achieved, $\mathbf{P}^L$ and $\mathbf{Z}_{l,k}^P$ are obtained. Similarly, the feature map $\mathbf{Z}_{l,k}^G\in\mathcal{R}^{h\times w}$ of $\mathbf{GS}_1^H$ and $\mathbf{GS}_1^L\in\mathcal{R}^{h\times w}$ of the first component $\mathbf{GS}_1$ also can be estimated according to (1).

B. Feature Map Fusion

For the fused result, the fusion rule on feature maps has a significant influence on spatial information. Because the spatial information in different scales has different statistical properties, the feature maps corresponding to filters with different size have great differences. Thus, the same fusion rule will lead to some spatial distortions in feature maps from different scales. Besides, the injection gain is generally calculated globally for the images to be fused, which ignores the local similarity among the pixels. Thus, some spectral distortions are introduced. In this section, taking the influence of different filter sizes on feature maps and local similarity into consideration, a more proper fusion rule is adopted, which is designed as

$$\mathbf{Z}_{l,k}^{F}(i,j) = (1-\mathbf{C}_{l,k}(i,j))\cdot\mathbf{Z}_{l,k}^{P}(i,j) + \mathbf{C}_{l,k}(i,j)\cdot\mathbf{Z}_{l,k}^{G}(i,j) \tag{11}$$

$$\mathbf{C}_{l,k}(i,j) = \frac{\sigma_{PG}}{\sigma_P\sigma_G}\cdot\frac{2\mu_P\mu_G}{\mu_P^2+\mu_G^2}\cdot\frac{2\sigma_P\sigma_G}{\sigma_P^2+\sigma_G^2} \tag{12}$$

where $\mathbf{Z}_{l,k}^{F}\in\mathcal{R}^{h\times w}$ is the fused feature map. (i,j) is the position of pixel. $\mathbf{C}_{l,k}(i,j)$ is the correlation between the patches $\mathbf{U}_{l,k}^{i,j}\in\mathcal{R}^{s_l\times s_l}$ and $\mathbf{V}_{l,k}^{i,j}\in\mathcal{R}^{s_l\times s_l}$ from $\mathbf{Z}_{l,k}^{P}$ and $\mathbf{Z}_{l,k}^{G}$, which is defined by universal image quality index (UIQI) [61]. Both patches $\mathbf{U}_{l,k}^{i,j}$ and $\mathbf{V}_{l,k}^{i,j}$ are centered on (i,j). For feature maps with larger filter size, the sparsity is more obvious than that with smaller filter size, so larger patch is needed to capture the local information. For convenience, the patch sizes of $\mathbf{U}_{l,k}^{i,j}$ and $\mathbf{V}_{l,k}^{i,j}$ are directly set as the size of filter $\mathbf{F}_{l,k}$ corresponding to the feature maps $\mathbf{Z}_{l,k}^{P}$ and $\mathbf{Z}_{l,k}^{G}$. For example, if the size of filter $\mathbf{F}_{l,k}$ is 3×3, then the sizes of $\mathbf{U}_{l,k}^{i,j}$ and $\mathbf{V}_{l,k}^{i,j}$ are set as 3×3. σ_{PG} is the covariance of the two patches. σ_P and μ_P are the standard variance and the mean of $\mathbf{U}_{l,k}^{i,j}$, respectively. Similarly, the standard variance and the mean of $\mathbf{V}_{l,k}^{i,j}$ are σ_G and μ_G. By (11), the statistical properties in different feature maps are fully considered and the spatial information in $\mathbf{Z}_{l,k}^{P}$ and $\mathbf{Z}_{l,k}^{G}$ is efficiently combined.

C. Low-Frequency Fusion

Generally, the extraction of low frequencies from $\mathbf{P}$ and $\mathbf{GS}_1$ has a significant influence on the fusion results. However, the separation of high and low frequencies is very difficult, especially PAN image containing abundant spatial details. Because the high and low frequencies cannot be separated efficiently and accurately in (1), low frequency $\mathbf{P}^L$ of $\mathbf{P}$ often contains some spatial details, which results in some spectral distortions and spatial blur effects in the fused images. Thus, a binary map $\mathbf{B}\in\mathcal{R}^{h\times w}$ is derived by combining the gradient information of $\mathbf{P}^L$ and $\mathbf{GS}_1^L$, which is calculated by

$$\mathbf{B}(i,j) = \begin{cases} 1, if \mathbf{M}^{\mathbf{P}}(i,j) > \mathbf{M}^{\mathbf{GS}}(i,j) \\ 0, \quad \text{otherwise} \end{cases} \tag{13}$$

$$\mathbf{M}^{\mathbf{P}}(i,j) = \sqrt{\nabla_h\mathbf{P}^L(i,j)^2+\nabla_v\mathbf{P}^L(i,j)^2} \tag{14}$$

$$\mathbf{M}^{\mathbf{GS}}(i,j) = \sqrt{\nabla_h\mathbf{GS}_1^L(i,j)^2+\nabla_v\mathbf{GS}_1^L(i,j)^2}. \tag{15}$$

Then, the fused low frequency $\mathbf{GS}_1^{LF}\in\mathcal{R}^{h\times w}$ is computed by

$$\mathbf{GS}_1^{LF} = \mathbf{B}\odot\mathbf{P}^L + (1-\mathbf{B})\odot\mathbf{GS}_1^L \tag{16}$$

where $\odot$ denotes elementwise multiplication. Because the regions with large gradient magnitude contain rich spatial information, the spatial details in $\mathbf{P}^L$ are also further injected into $\mathbf{GS}_1^L$ through (16), which can better enhance the spatial details and preserve the spectral information in the fused image.

D. Fused Image Reconstruction

When the feature maps of $\mathbf{P}^H$ and $\mathbf{GS}_1^H$ are fused by (11), the fused high frequency $\mathbf{GS}_1^{HF}$ can be reconstructed recursively by

$$\mathbf{GS}_1^{HF} = \sum_{l=1}^{L}\sum_{k=1}^{K_l}\mathbf{F}_{l,k}*\mathbf{Z}_{l,k}^{F}. \tag{17}$$

Then, the fused first component $\mathbf{GS}_1^F$ can be computed by

$$\mathbf{GS}_1^F = \mathbf{GS}_1^{LF} + \mathbf{GS}_1^{HF}. \tag{18}$$

Finally, the fused HR MS image is obtained by the inverse GS transform on $\mathbf{GS}_1^F$ and other GS components.

III. Details for MCSD

A. CSD

CSD [46]–[47] assumes that a given image $\mathbf{E}\in\mathcal{R}^{h\times w}$ can be decomposed into a series of feature maps $\{\mathbf{Z}_k\}_{k=1,2,\ldots,K}\in\mathcal{R}^{h\times w}$ with corresponding filters $\{\mathbf{F}_k\}_{k=1,2,\ldots,K}\in\mathcal{R}^{s\times s}$. The size of filter $\mathbf{F}_k$ is $s\times s$. K is the number of filters or feature maps. Similar to SR [23], sparsity is imposed on feature maps for a reasonable solution. Then, CSD is formulated as

$$\underset{\{\mathbf{Z}_k\}}{arg\min}\frac{1}{2}\left\|\mathbf{E}-\sum_{k=1}^{K}\mathbf{F}_k*\mathbf{Z}_k\right\|_2^2+\lambda\sum_{k=1}^{K}\|\mathbf{Z}_k\|_1 \tag{19}$$

where λ is a tradeoff parameter and $\|\cdot\|_1$ denotes the sum of absolute values for all elements in $\mathbf{Z}_k$. Due to avoiding the image partition, CSD can locally form a more efficient representation for a global image compared with SR [53]. Thus, CSD can achieve more accurate image reconstruction than SR. However, it is difficult to optimize (1) because of the involved convolution. Zeiler *et al*. [46] designed an optimization approach in spatial domain, whose computation complexity is non-negligible and reconstruction performance is limited. Subsequently, some optimization algorithms [54] are introduced by ADMM approach, which solve a large linear system after FT.

B. MCSD

Obviously, the filters with the same size in (19) are employed to reconstruct the image. So it is difficult to well represent the multiscale details by single-scale filters, as demonstrated in Fig. 1. In order to improve the representation capacity, multiscale filters are considered to achieve the decomposition of the image. Then, the MCSD model can be formulated as

$$\mathbf{E} = \sum_{l=1}^{L}\sum_{k=1}^{K_l}\mathbf{F}_{l,k}*\mathbf{Z}_{l,k}+\mathbf{N} \tag{20}$$

where $\mathbf{Z}_{l,k}\in\mathcal{R}^{h\times w}$ is the *k*th feature map in the *l*th scale. $\mathbf{N}\in\mathcal{R}^{h\times w}$ is the additive white Gaussian noise. Then, sparsity is imposed on feature maps to regularize the solution space, whose formulation can be molded as

$$\underset{\{\mathbf{Z}_{l,k}\}}{arg\min}\frac{1}{2}\left\|\mathbf{E}-\sum_{l=1}^{L}\sum_{k=1}^{K_l}\mathbf{F}_{l,k}*\mathbf{Z}_{l,k}\right\|_2^2+\lambda\sum_{l=1}^{L}\sum_{k=1}^{K_l}\|\mathbf{Z}_{l,k}\|_1 \tag{21}$$

where λ is a tradeoff parameter. Considering the coupling of $\mathbf{Z}_{l,k}$, an auxiliary variable $\mathbf{X}_{l,k}\in\mathcal{R}^{h\times w}$ associating with $\mathbf{Z}_{l,k}$

is introduced to optimize (21) alternatively. Then, (21) can be rewritten as

$$\underset{\{\mathbf{Z}_{l,k}\}}{arg\min}\ \frac{1}{2}\left\|\mathbf{E}-\sum_{l=1}^{L}\sum_{k=1}^{K_l}\mathbf{F}_{l,k}*\mathbf{X}_{l,k}\right\|_2^2+\lambda\sum_{l=1}^{L}\sum_{k=1}^{K_l}\|\mathbf{Z}_{l,k}\|_1.$$
$$\text{s.t.}\mathbf{X}_{l,k}=\mathbf{Z}_{l,k} \tag{22}$$

ADMM is derived to solve (22) and the augmented Lagrange function can be written as

$$F=\frac{1}{2}\left\|\mathbf{E}-\sum_{l=1}^{L}\sum_{k=1}^{K_l}\mathbf{F}_{l,k}*\mathbf{X}_{l,k}\right\|_2^2+\lambda\sum_{l=1}^{L}\sum_{k=1}^{K_l}\|\mathbf{Z}_{l,k}\|_1$$
$$+\frac{\mu}{2}\sum_{l=1}^{L}\sum_{k=1}^{K_l}\|\mathbf{X}_{l,k}-\mathbf{Z}_{l,k}\|_2^2+\mu\sum_{l=1}^{L}\sum_{k=1}^{K_l}\langle\mathbf{Y}_{l,k},\mathbf{X}_{l,k}-\mathbf{Z}_{l,k}\rangle \tag{23}$$

where $\mathbf{Y}_{l,k}\in\mathcal{R}^{h\times w}$ is the Lagrange multiplier and μ is a penalty parameter. According to the framework of ADMM, $\mathbf{X}_{l,k}$, $\mathbf{Z}_{l,k}$, and $\mathbf{Y}_{l,k}$ are alternatively updated. For $\mathbf{Z}_{l,k}$ and $\mathbf{X}_{l,k}$, similar solutions with (3) and (6) can be derived.

Finally, the multiplier is computed by

$$\mathbf{Y}_{l,k}=\mathbf{Y}_{l,k}+\mu\left(\mathbf{X}_{l,k}-\mathbf{Z}_{l,k}\right). \tag{24}$$

By alternative and iterative update above, the feature maps corresponding to multiscale filters can be estimated efficiently.

In (21), it is assumed that the multiscale filters are known. In order to learn more compact and representative filters, a multiscale filter learning model is established. For N training images, the multiscale filter learning task can be defined as

$$\underset{\{\mathbf{F}_{l,k}\},\{\mathbf{Z}_{l,k}^n\}}{arg\min}\ \frac{1}{2}\sum_{n=1}^{N}\left\|\mathbf{E}^n-\sum_{l=1}^{L}\sum_{k=1}^{K_l}\mathbf{F}_{l,k}\right.$$
$$\left.*\mathbf{Z}_{l,k}^n\right\|_2^2+\gamma\sum_{n=1}^{N}\sum_{l=1}^{L}\sum_{k=1}^{K_l}\left\|\mathbf{Z}_{l,k}^n\right\|_1$$
$$\text{s.t. }\|\mathbf{F}_{l,k}\|_2=1\forall l,\forall k \tag{25}$$

where γ is a weighting parameter. In (25), filters and feature maps are alternatively calculated, whose detailed optimization algorithm can be found in [47]. By (25), some essential and intrinsic multiscale filters can be learned from many images, which are then adopted by MCSD.

IV. Experimental Results and Comparisons

In this section, experimental settings are introduced in details. Then, experimental results on different datasets are presented and analyzed, which demonstrate the performance of the proposed method. Finally, parameter analysis and running time comparison are comprehensively investigated for the proposed method.

A. Experiment Setup

In the following parts, the fusion experiments are conducted on four pairs of 64×64 LR MS image and 256×256 PAN image from QuickBird and GeoEye-1 satellites, which are displayed in Fig. 3. QuickBird satellite can produce PAN image and LR MS image, whose corresponding resolutions are 0.61 and 2.44 m at nadir. For GeoEye-1 satellite, the resolutions of PAN and LR MS images are 0.41 and 1.64 m at nadir, respectively. The image pairs in Fig. 3(a)–(b) and (e)–(f) are captured from Sundarbans, India on November 21, 2002 by QuickBird satellite. The reduced-resolution PAN and LR MS images in Fig. 3(a) and (b) are produced by blurring and downsampling with rate 4, and then the fused image of them is compared with the original MS image, also named as reference image. For reduced-resolution dataset, blurring operation is achieved by MTF filter, whose frequency response is approximately Gaussian shaped. The shapes of filters for different bands are slightly different and their gains at Nyquist cutoff frequency can be found in [56]. GeoEye-1 satellite provides the reduced-resolution images in Fig. 3(c) and (d) and the full-resolution images in Fig. 3(g) and (h). The reduced-resolution images in Fig. 3(c) and (d) are generated in the same way as that for the images in Fig. 3(a) and (b) and the fusion result is compared with their corresponding reference image. In order to analyze the fusion results of the proposed method, some related methods, such as variational pansharpening with local gradient constraints (VPLGC) [57], proportional additive wavelet LHS (AWLP) [15], SVT [20], two-step sparse coding with patch normalization (PN-TSSC) [58] are employed. Besides, some methods based on CSD, such as convolution structure sparse coding (CSSC) [48], convolutional sparse representation fusion (CSRF) [49], and convolutional sparse representation of injected details (CSR-D) [50] are also considered for visual and numerical evaluations. Moreover, CNN-based method in [39] and DRPNN [43] are also compared. In reduced-resolution cases, four indexes, Q4 [59], spectral angle mapper (SAM) [60], UIQI [61], and *Erreur Relative Globale Adimensionnelle de Synthèse* (ERGAS) [62] are utilized for quantitative evaluation. D_λ, D_S, and quality no reference (QNR) proposed in [63] are also employed for the assessment of the full-resolution images because there are no reference images for comparison.

B. Parameter Settings

In the proposed method, the filter number and scale number are 12 and 3, respectively. The three scales are 3×3, 7×7, and 11×11, respectively. Each scale involves four filters. For the tradeoff parameters, α and β are set as 2^5 and 1, respectively. The maximum iteration number is set as 200 to solve (1). The minimum relative reconstruction error is 10^{-5}. The elements of variables to be solved in (1) are initialized by 0. Moreover, the penalty parameter μ is set as 0.5 in the first iteration and then increases by multiplying a small gain $\rho=1.3$ in each iteration. Besides, the multiscale filters are trained by SPORCO toolbox [47] in which 50 full-resolution PAN images with size 256×256 are used as training images and they are collected from QuickBird and GeoEye-1 satellites. In order to obtain finer filters, the low frequencies of the training images are subtracted, which are produced by convolving training images to a 9×9 Gaussian kernel with standard derivation 10. Then, the

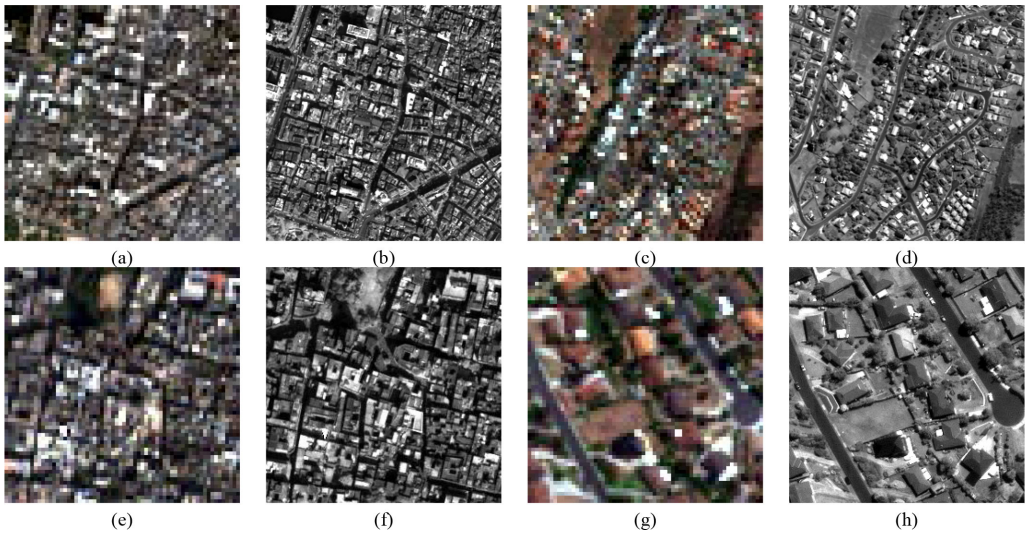

Fig. 3. Image pairs to be fused. (a) Reduced-resolution LR MS image 1. (b) Reduced-resolution PAN image 1. (c) Reduced-resolution LR MS image 2. (d) Reduced-resolution PAN image 2. (e) Full-resolution LR MS image 1. (f) Full-resolution PAN image 1. (g) Full-resolution LR MS image 2. (h) Full-resolution PAN image 2.

corresponding high frequencies are employed for filter training. For the sparsity constraint of feature maps, γ is 0.5 for training data.

C. Experiments on Reduced-Resolution Dataset

In this part, the experiments are conducted on two pairs of reduced-resolution images, which are displayed in Fig. 3(a)–(d). The fusion results of all methods are illustrated in Fig. 4 and some local regions are selected and amplified for direct analysis, which are circled by a red rectangle. Besides, the error maps between fused images and reference image are also displayed. The first image in Fig. 4(a) is the reference image and then the fused images are directly compared with the reference image. The result in Fig. 4(b) is from VPLGC [57] and we can see that some spatial blurring effects appear especially in building regions. Although the spatial details are enhanced well, spectral distortions can be found in Fig. 4(c) and (d). The result of SVT [20] in Fig. 4(e) preserves the spectral features. In Fig. 4(f), the result also suffers from the spectral distortions. On the contrary, some blurring effects arise in the upper-right regions in Fig. 4(g), which may be caused by improper filters. For CSRF [49], the color of trees looks unnatural. In Fig. 4(i), the spatial details are sharp but some spectral differences can be found. For the proposed method, it can be found that the spectral features in Fig. 4(k) are more consistent with those of the reference image. Besides, we can see that the spatial details in local regions are very close to the reference image for different methods. But some spectral differences appear. For example, the local region in the result of PNN [39] is more colorful than those in the other results. However, the color in the local region of PNN [39] looks unnatural when compared with the reference local region. Besides, we also illustrate the error maps of the fused images from all methods and the red band is selected for comparison. From the error maps, it can be observed that the reconstructed errors of VPLGC [57], CSR-D [50], and PNN [39] are considerable. The error of the proposed method is smaller when compared with other methods visually.

The quantitative values of all fused results in Fig. 4 are listed in Table I, where the best result for each index is labeled in bold. It can be observed that the best values of SAM, UIQI, and ERGAS are generated by the proposed method MCSD. ERGAS measures the whole spectral distortions in the fused images. So the best ERGAS for Fig. 4(k) means that MCSD performs well in spectral preservation. Besides, SVT [20] provides the best Q4. But the difference between Q4 of MCSD and that of SVT [20] is small.

Fig. 5 shows the results of all methods on reduced-resolution GeoEye-1 satellite dataset, where the reference image is given in Fig. 5(a). A local region is also chosen and enlarged for comparison, whose location is labeled by a red rectangle. Similar to the result in Fig. 4, the result of VPLGC [57] also contains some bluring effects when compared with the reference image in Fig. 5(a). Some spectral distortions can be found in the result of DRPNN [43]. AWLP [15] and SVT [20] belong to the same category, but the result of SVT [20] has a better performance than that of AWLP [15] visually. For PN-TSSC [58], there are great color differences in the building regions compared with Fig. 5(a). The spatial information is enhanced well in the results of CSSC [48] and CSR-D [50], but the spectral information is distorted in Fig. 5(i). The fused image in Fig. 5(h) lacks many spatial edges or textures in some buildings. The proposed method MCSD behaves well in spatial details and the spectral information is also preserved better than the other fused images. From the enlarged areas, it can be found there are some spectral differences on the roof for different methods when compared with the reference image. The color of local area in PNN [39] is closer to that of reference image. However, its reconstruction error is obvious. Besides, we can see spatial details appear in some error maps, such as CSRF [49] and PNN [39]. In addition, we also present

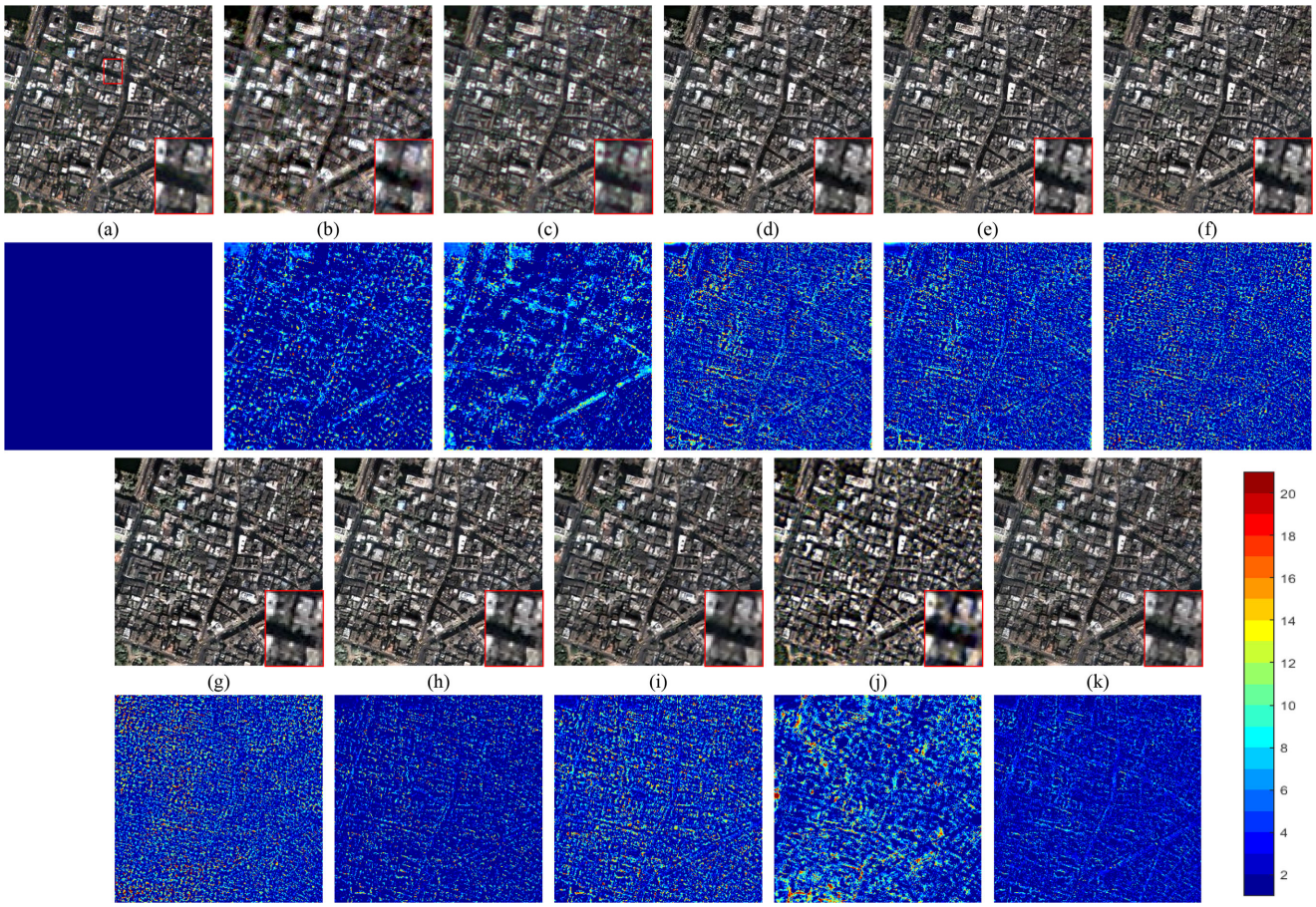

Fig. 4. Fusion results and error maps of images from QuickBird. The error maps in the second row and fourth row correspond the images in the first row and third row, respectively. (a) Reference image. (b) VPLGC [57]. (c) DRPNN [43]. (d) AWLP [15]. (e) SVT [20]. (f) PN-TSSC [58]. (g) CSSC [48]. (h) CSRF [49]. (i) CSR-D [50]. (j) PNN [39]. (k) MCSD.

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

Metric	VPLGC[57]	DRPNN[43]	AWLP[15]	SVT[20]	PN-TSSC[58]	CSSC[48]	CSRF[49]	CSR-D[50]	PNN[39]	MCSD
Q4	0.8298	0.8553	0.8840	**0.8893**	0.8768	0.8887	0.8718	0.8891	0.8558	0.8831
SAM	9.9560	9.0824	10.1127	9.3551	9.9039	9.0256	8.5531	8.9538	14.9718	**8.4011**
UIQI	0.9054	0.9096	0.9180	0.9238	0.9325	0.9179	0.9023	0.9268	0.9225	**0.9332**
ERGAS	3.1628	2.9797	3.2040	2.9947	2.8344	3.0647	3.5490	2.8233	3.8884	**2.7408**

the error maps of all methods in Fig. 5. In the error map of the proposed method, some differences in low frequency can be observed, which may be caused by the excessive information introduction of PAN image through low frequency fusion rule.

Table II lists the quality assessment results of all methods. We can see that PN-TSSC [58] provides the best Q4. However, the proposed method MCSD gives the best values for SAM, UIQI, and ERGAS. Therefore, the overall quality of MCSD is better than other methods in numerical comparison.

D. Experiments on Full-Resolution Dataset

In this experiment, QuickBird and GeoEye-1 satellites provide two pairs of real images for qualitative and quantitative evaluation. Fig. 6 reports the fusion results of all methods on QuickBird dataset and an interesting area circled by red line is selected for analysis. For a full-resolution dataset, there is no reference image and error maps for direct comparison. It can be observed that VPLGC [57] can provide better spectral information but spatial details are smoothed in the fused image. In the fusion results of AWLP [15] and SVT [20], the spectral information is preserved better. Some spatial overlapping effects are visible in the result of PN-TSSC [58], which may be caused by image partition. In Fig. 6(f), the spatial information is enhanced well because CSSC [48] models the image globally. However, obvious spectral distortion can be seen in the result of CSRF [49]. For CSR-D [50], we can see some details in building regions are blurred. In Fig. 6(h), the spectral features for the building in the top right corner are distorted. However,

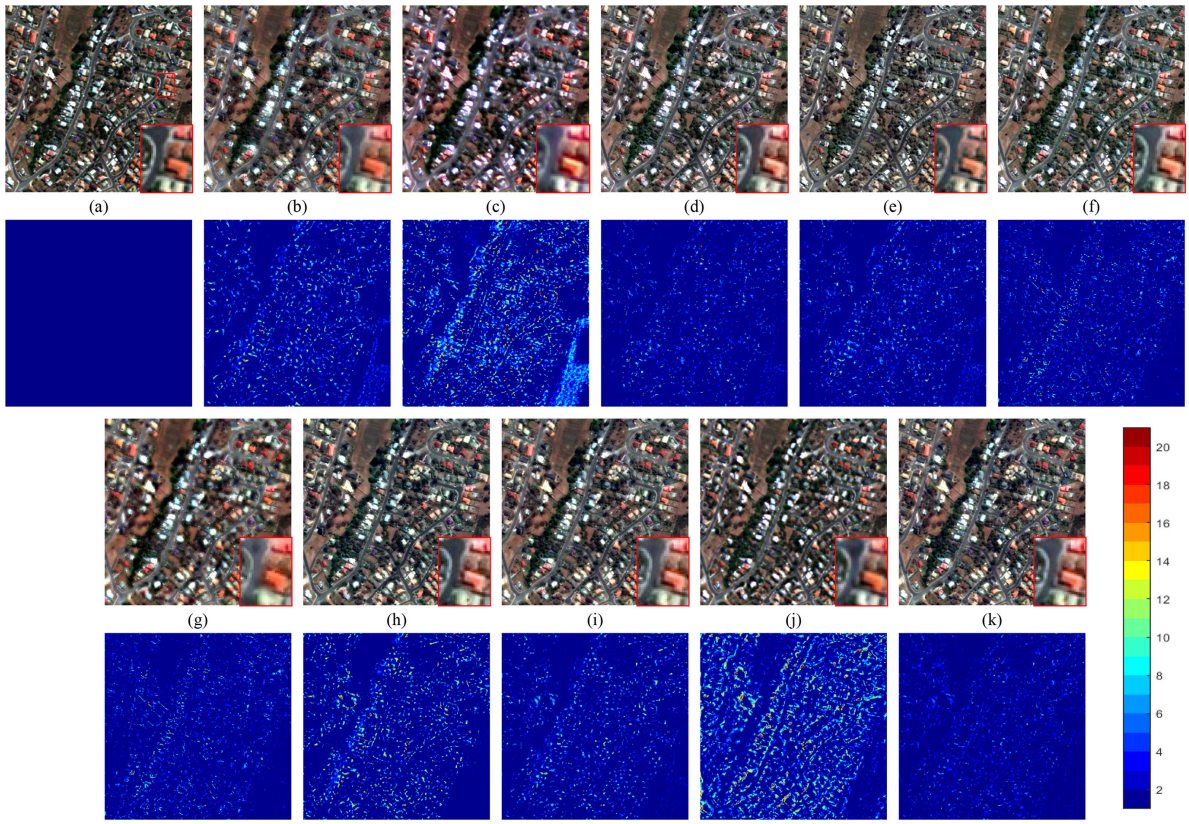

Fig. 5. Fusion results and error maps of images from GeoEye-1. The error maps in the second row and fourth row correspond the images in the first row and third row, respectively. (a) Reference image. (b) VPLGC [57]. (c) DRPNN [43]. (d) AWLP [15]. (e) SVT [20]. (f) PN-TSSC [58]. (g) CSSC [48]. (h) CSRF [49]. (i) CSR-D [50]. (j) PNN [39]. (k) MCSD.

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

Metric	VPLGC[57]	DRPNN[43]	AWLP[15]	SVT[20]	PN-TSSC[58]	CSSC[48]	CSRF[49]	CSR-D[50]	PNN[39]	MCSD
Q4	0.7446	0.7685	0.7708	0.7578	**0.7849**	0.7231	0.6392	0.7781	0.7391	0.7736
SAM	5.8566	6.5016	6.1567	6.0838	6.2902	6.6617	6.4030	6.3643	7.4354	**5.6867**
UIQI	0.9210	0.9197	0.9228	0.9140	0.9265	0.8861	0.8816	0.9282	0.9114	**0.9354**
ERGAS	1.7311	1.9032	1.7176	1.8033	3.4662	3.2591	3.4788	3.4369	2.1631	**1.6743**

the color in the same position is maintained better in Fig. 6(j). Moreover, the spatial details in Fig. 6(j) look clearer than other fused images. From the amplified local region, some spectral distortions arise in Fig. 6(d) and (i) and blurring effects can be seen in Fig. 6(g). The result in Fig. 6(j) looks better.

Table III provides the values of D_S, D_λ, and QNR for numerical comparison, in which the best results are labeled in bold. It can be found that the numerical values are consistent with the visual analysis in Fig. 6. PNN [39] provides the best D_s. However, the best values of D_λ and QNR are from the proposed method MCSD.

Besides, Fig. 7 displays the fusion images of all methods on GeoEye-1 dataset. Meanwhile, some local regions are also enlarged and put in the bottom right corner of the fused image for comparison. In Fig. 7, we can also see that the spatial effects appear in Fig. 7(a). The result in Fig. 7(b) performs better in spectral signatures. Fig. 7(d) has a similar visual performance with Fig. 7(f), in which some regions containing vegetation become dark-grey. In Fig. 7(g), the spatial information loss can be found and there are some spectral differences among them. For MCSD, the spatial details in Fig. 7(j) are enhanced well. In the local region of PNN [39], some spectral artifacts can be seen. The details in the local region of Fig. 7(g) are blurry. In the selected area, the color in Fig. 7(j) is natural.

The numerical values of all indexes are calculated and listed in Table IV. The best D_s is from CSRF [49], which is followed by the proposed method. VPLGC [57] provides the best D_λ. As an overall evaluation index, the best QNR is from the proposed method and the QNR of PNN [39] is the second-best.

E. Investigation on Multiscale Filters

In this part, we investigate the performance of the proposed method with different multiscale filters on reduced-resolution GeoEye-1 dataset, which is shown in Fig. 3(c) and (d). In order to analyze the influences of the number of scales, we consider

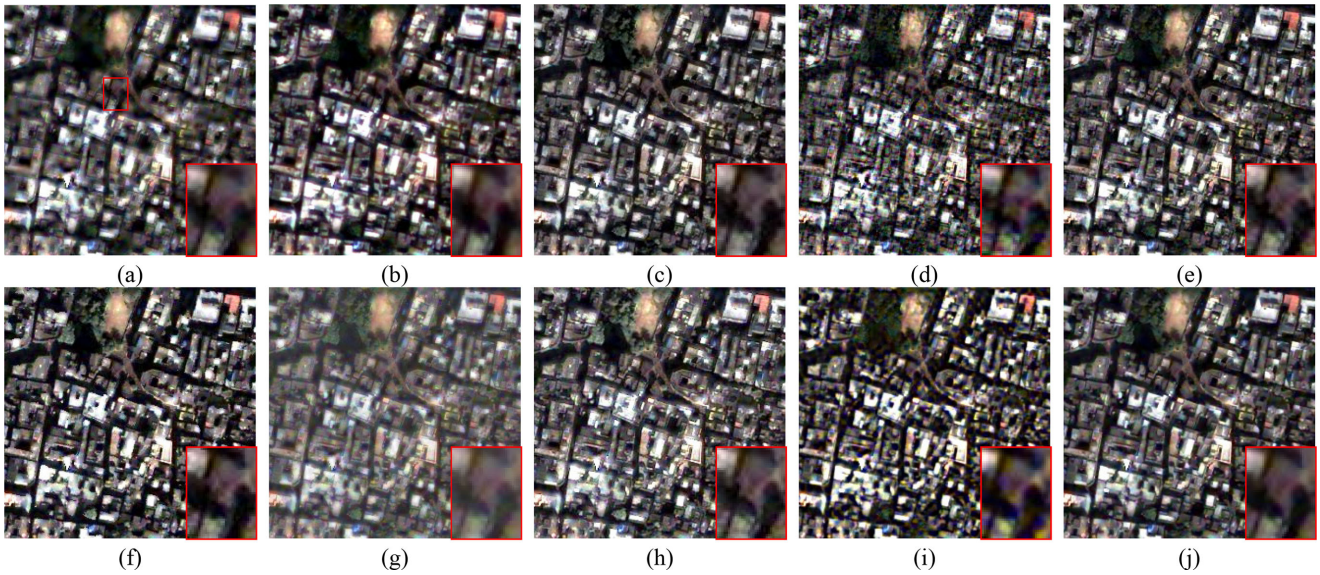

Fig. 6. Fusion results of images from QuickBird. (a) VPLGC [57]. (b) DRPNN [43]. (c) AWLP [15]. (d) SVT [20]. (e) PN-TSSC [58]. (f) CSSC [48]. (g) CSRF [49]. (h) CSR-D [50]. (i) PNN [39]. (j) MCSD.

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

Metric	VPLGC[57]	DRPNN[43]	AWLP[15]	SVT[20]	PN-TSSC[58]	CSSC[48]	CSRF[49]	CSR-D[50]	PNN[39]	MCSD
D_λ	0.0299	0.0380	0.0669	0.0659	0.0570	0.0647	0.0336	0.0380	0.1123	**0.0254**
D_s	0.0902	0.0646	0.0777	0.0843	0.0582	0.0505	0.0944	0.0478	**0.0381**	0.0385
QNR	0.8826	0.8998	0.8606	0.8562	0.8882	0.8880	0.8752	0.9160	0.8539	**0.9371**

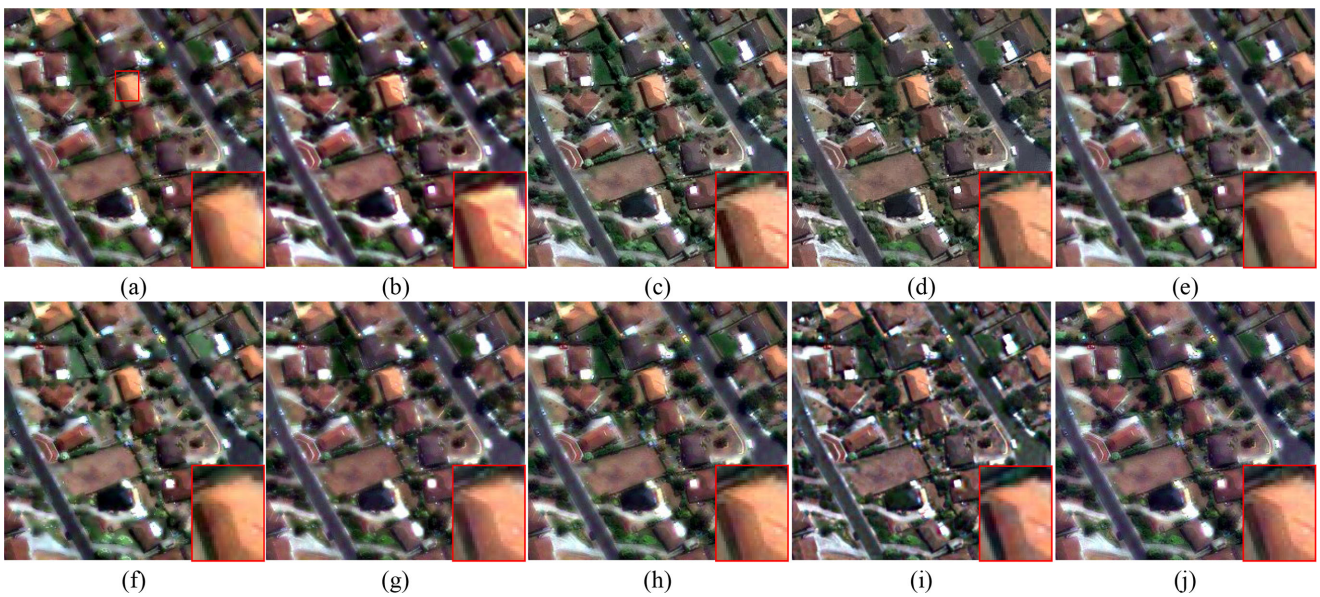

Fig. 7. Fusion results of images from GeoEye-1. (a) VPLGC [57]. (b) DRPNN [43]. (c) AWLP [15]. (d) SVT [20]. (e) PN-TSSC [58]. (f) CSSC [48]. (g) CSRF [49]. (h) CSR-D [50]. (i) PNN [39]. (j) MCSD.

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

Metric	VPLGC[57]	DRPNN[43]	AWLP[15]	SVT[20]	PN-TSSC[58]	CSSC[48]	CSRF[49]	CSR-D[50]	PNN[39]	MCSD
D_λ	**0.0532**	0.0613	0.1136	0.1173	0.0989	0.0561	0.0989	0.0770	0.0346	0.0624
D_s	0.0878	0.0906	0.0702	0.0695	0.0518	0.0753	**0.0330**	0.0445	0.0734	0.0432
QNR	0.8637	0.8537	0.8242	0.8214	0.8544	0.8729	0.8713	0.8818	0.8945	**0.8971**

TABLE V
NUMERICAL EVALUATION OF FUSED RESULTS WITH DIFFERENT SCALES OF FILTERS

Metric	7×7×12	7×7×6 11×11×6	3×3×4 7×7×4 11×11×4	3×3×3 7×7×3 11×11×3 15×15×3
Q4	**0.7761**	0.7757	0.7736	0.7759
SAM	5.6988	5.6890	**5.6867**	5.6895
UIQI	0.9266	0.9299	**0.9354**	0.9296
ERGAS	1.8317	1.7738	**1.6743**	1.7797

TABLE VI
NUMERICAL EVALUATION OF FUSED RESULTS WITH DIFFERENT NUMBER OF FILTERS

Metric	3×3×1 7×7×1 11×11×1	3×3×2 7×7×2 11×11×2	3×3×4 7×7×4 11×11×4	3×3×8 7×7×8 11×11×8	3×3×16 7×7×16 11×11×16
Q4	0.7688	0.7722	0.7736	0.7742	**0.7748**
SAM	**5.6834**	5.6861	5.6867	5.6870	5.6883
UIQI	0.9333	0.9352	**0.9354**	0.9351	0.9351
ERGAS	1.6992	1.6754	**1.6743**	1.6811	1.6815

different combinations with different filter sizes, as shown in the first line in Table V. For different combinations, the total number of filters is 12. With the same number of filters, the filter size varies from 3×3 to 15×15. From Table V, it can be observed that the numerical values vary with different scale filters. For single scale $7 \times 7 \times 12$, the fusion result has a poor performance in SAM, UIQI, and ERGAS. Although the best Q4 is from the single scale, the fusion result from three scales behaves better in the other three indexes. In Table V, the proposed method with four scales cannot produce satisfactory results, probably because the filters with larger size cannot efficiently exploit the spatial details of the images in Fig. 3(c) and (d).

Besides, the influences of number of filters on the fusion results are also conducted for a more comprehensive analysis. The total number of all filters increases from 3 to 48. The detailed settings about filters are listed in the first line of Table VI. We can find that the best values of UIQI and ERGAS are achieved when the number of filters is 12. Although the best SAM is produced by the setting with three filters, the other indexes are poor. Moreover, with the increasing of the number of filters, the implementation time also increases because more feature maps need to be estimated. Thus, the setting with 12 filters is utilized in the proposed method.

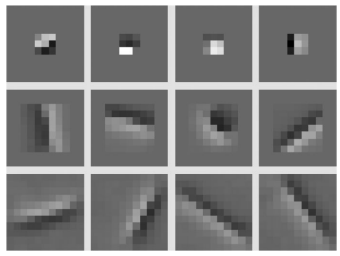

Fig. 8. Multiscale filters learned by the proposed method.

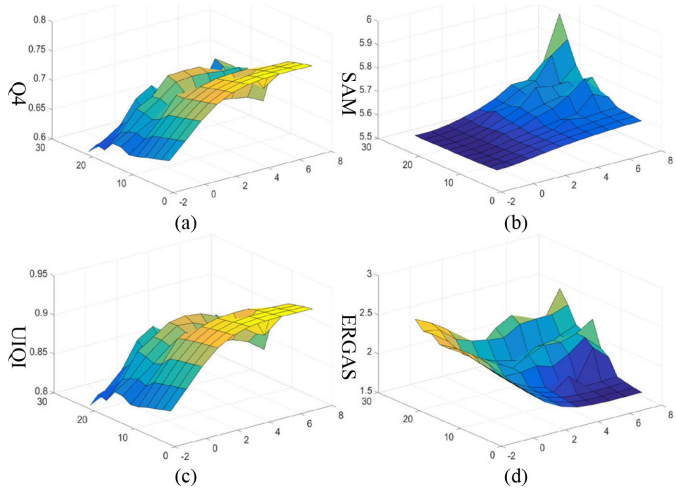

Fig. 9. Influences of α and β on fusion result of the proposed method on reduced-resolution GeoEye-1 dataset.

With the same setting in the third column of Table VI, the learned filters by the proposed method are demonstrated in Fig. 8. In the first line of Fig. 8, the filter size is 3×3. The filter sizes in the second and last lines are 7×7 and 11×11, respectively. In Fig. 8, filters with different orientations can effectively approximate the spatial textures and the filters with smaller size are used to model finer spatial details. Therefore, the images to be fused can be decomposed well by these filters.

F. Investigation on Parameters

In this part, we analyze the influences of parameters α and β on the fusion results of the image pair in Fig. 3(c) and (d). α varies from 2^{-2} to 2^7 with step size 1 on the power. Besides, β increases from 1 to 21 with step size 2. In Fig. 9, the coordinates in the horizontal and vertical axises equal the log base 2 of α and β, respectively. We can see that better Q4, UIQI, and ERGAS can be obtained with increasing α. But the result of SAM is worse. Besides, satisfactory values are produced with smaller

TABLE VII
TIME COMPARISON OF ALL METHODS (S)

Method	Time
VPLGC [57]	7.10
DRPNN [43]	3.08
AWLP [15]	0.53
SVT [20]	0.86
PN-TSSC [58]	4.95
CSSC [48]	933.68
CSRF [49]	18.11
CSR-D [50]	441.47
PNN [39]	4.85
MCSD	30.92

β. Through comprehensive consideration, α and β are finally set as 2^5 and 1.

G. Time Analysis

In this part, we analyze the running time of all methods on 64×64 LR MS and 256×256 PAN image pair. The experiments are performed by MATLAB R2017a on the same device with Core i7-6700/3.4GHz/16G. Table VII lists the runtime of different methods, in which the time is measured in second. From Table VII, it can be found that CSSC [48] is the most time-consuming method, because more filters, about 70 filters, are used to represent the images to be fused. Besides, CSR-D [50] also spends a lot of time to obtain the fusion result. In CSR-D [50], the same decomposition procedure about CSD is implemented four times because MS images are made up of four bands. Compared with the methods based on CS or MRA, CSD based methods are time-consuming. The proposed method is more efficient when compared with CSSC [48] and CSR-D [50] due to rapid convergence and fewer filters in MCSD.

V. CONCLUSION

In this article, we proposed an efficient pansharpening method based on MCSD. For LR MS and PAN images, the multiscale feature maps and low frequencies of them are simultaneously obtained by a unified MCSD framework, which can better model the multiscale spatial and spectral information in LR MS and PAN images. For the feature maps from LR MS and PAN images, the fusion rule is designed by calculating the local similarity among them, which further exploits the multiscale information in feature maps. Besides, in order to take full advantage of the spatial information in the low frequency of PAN image, a binary map is estimated by comparing the gradient magnitudes to fuse the low frequencies because the regions containing abundant spatial details generally have larger gradient magnitudes. Finally, the fused image is generated after reconstruction of fused low frequencies and feature maps. Compared with VPLGC [57], DRPNN [43], AWLP [15], SVT [20], PN-TSSC [58], CSSC [48], CSRF [49], CSR-D [50], and PNN [39], the proposed method demonstrates a better performance in objective and subjective evaluations. Due to the introduction of multiscale property, the proposed method can effectively enhance the spatial details and preserve the spectral information in the fused images. For future work, more proper transform will be employed or reasonable fusion rules are designed to obtain better fusion results.

REFERENCES

[1] G. Cheng, P. Zhou, and J. Han, "Learning rotation invariant convolutional neural networks for object detection in VHR optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7405–7415, Dec. 2016.

[2] X. Liu *et al.*, "Deep multiple instance learning-based spatial-spectral classification for PAN and MS imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 461–473, Jan. 2018.

[3] W. Zhang, X. Lu, and X. Li, "A coarse-to-fine semi-supervised change detection for multispectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3587–3599, Jun. 2018.

[4] H. Ghassemian, "A review of remote sensing image fusion methods," *Inf. Fusion*, vol. 32, pp. 75–89, Nov. 2016.

[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] R. Dian, S. Li, and L. Fang, "Learning a low tensor-train rank representation for hyperspectral image super-resolution," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2672–2683, Sep. 2019.

[7] 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.

[8] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.

[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 US6011875A, Jan. 2000.

[11] Y. Kim, M. Kim, J. Choi, and Y. Kim, "Image fusion of spectrally nonoverlapping imagery using SPCA and MTF-based filters," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2295–2299, Dec. 2017.

[12] G. Vivone, "Robust band-dependent spatial-detail approaches for panchromatic sharpening," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 57, no. 9, pp. 6421–6433, Sep. 2019.

[13] J. Duran and A. Buades, "Restoration of pansharpened images by conditional filtering in the PCA domain," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 3, pp. 442–446, Mar. 2019.

[14] 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.

[15] X. Otazu, M. González-Audícana, 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.

[16] L. Alparone, S. Baronti, B. Aiazzi, and A. Garzelli, "Spatial methods for multispectral pansharpening: Multiresolution analysis demystified," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 2563–2576, May 2016.

[17] G. Vivone, R. Restaino, and J. Chanussot, "A regression-based high-pass modulation pansharpening approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 984–996, Feb. 2018.

[18] R. Restaino, G. Vivone, P. Addesso, and J. Chanussot, "A pansharpening approach based on multiple linear regression estimation of injection coefficients," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 1, pp. 102–106, Jan. 2020.

[19] G. Vivone, S. Marano, and J. Chanussot, "Pansharpening: Context-based generalized laplacian pyramids by robust regression," *IEEE Trans. Geosci. Remote Sens*, vol. 58, no. 9, pp. 6152–6167, Sep. 2020.

[20] 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.

[21] S. Yang, M. Wang, and L. Jiao, "Fusion of multispectral and panchromatic images based on support value transform and adaptive principal component analysis," *Inf. Fusion*, vol. 13, no. 3, pp. 177–184, Jul. 2012.

[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] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[24] 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.
[25] M. R. Vicinanza, R. Restaino, G. Vivone, M. D. Mura, and J. Chanussot, "A pansharpening method based on the sparse representation of injected details," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 180–184, Jan. 2015.
[26] 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. 2015.
[27] 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. Top. Appl. Earth Observ. Remote Sens.*, vol. 9, no. 12, pp. 5740–5747, Dec. 2016.
[28] 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.
[29] S. Mei, J. Hou, J. Chen, L. Chau, and Q. Du, "Simultaneous spatial and spectral low-rank representation of hyperspectral images for classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2872–2886, May 2018.
[30] X. Dong *et al.*, "Weighted locality collaborative representation based on sparse subspace," *J. Vis. Commun. Image Represent.*, vol. 58, pp. 187–194, Jan. 2019.
[31] Y. Fang, K. Zhang, and Y. Ren, "Graph regularised sparse NMF factorisation for imagery de-noising," *IET Comput. Vis.*, vol. 12, no. 4, pp. 466–475, May 2018.
[32] D. Zeng *et al.*, "Pan-sharpening with structural consistency and $\ell 1/2$ gradient prior," *Geosci. Remote Sens. Lett.*, vol. 7, no. 12, pp. 1170–1179, 2016.
[33] S. Yang, J. Hou, Y. Jia, S. Mei, and Q. Du, "Hyperspectral image classification via sparse representation with incremental dictionaries," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 9, pp. 1598–1602, Sep. 2020.
[34] C. Chen, Y. Meng, Q. Luo, and Z. Zhou, "A novel variational model for pan-sharpening based on L1 regularization," *Geosci. Remote Sens. Lett.*, vol. 9, no. 2, pp. 170–179, 2018.
[35] S. Yang, J. Hou, Y. Jia, S. Mei, and Q. Du, "Pseudo-label guided kernel learning for hyperspectral image classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 12, no. 3, pp. 1000–1011, Mar. 2019.
[36] S. Mei, J. Ji, J. Hou, X. Li, and Q. Du, "Learning sensor-specific spatial-spectral features of hyperspectral images via convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4520–4533, Aug. 2017.
[37] Z. Zhu, J. Hou, J. Chen, H. Zeng, and J. Zhou, "Hyperspectral image super-resolution via deep progressive zero-centric residual learning," *IEEE Trans. Image Process.*, to be published.
[38] 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.
[39] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, 2016, Art. no. 594.
[40] G. Scarpa, S. Vitale, and D. Cozzolino, "Target-adaptive CNN-based pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5443–5457, Sep. 2018.
[41] Z. Shao and J. Cai, "Remote sensing image fusion with deep convolutional neural network," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1656–1669, May 2018.
[42] X. Fu, W. Wang, Y. Huang, X. Ding, and J. Paisley, "Deep multiscale detail networks for multiband spectral image sharpening," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published.
[43] Y. Wei, Q. Yuan, H. Shen, and L. Zhang, "Boosting the accuracy of multispectral image pansharpening by learning a deep residual network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1795–1799, Oct. 2017.
[44] Y. Wei and Q. Yuan, "Deep residual learning for remote sensed imagery pansharpening," in *Proc. IEEE Int. Workshop Remote Sens. Intell. Process.*, 2017, pp. 1–4.
[45] D. Wang, Y. Li, L. Ma, Z. Bai, and J. Chan, "Going deeper with dense connectedly convolutional neural networks for multispectral pansharpening," *Remote Sens.*, vol. 11, no. 22, p. 2608, Nov. 2019.
[46] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Aug. 2010, pp. 2528–2535.
[47] B. Wohlberg, "Efficient algorithms for convolutional sparse representations," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 301–315, Jan. 2016.
[48] 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.
[49] H. Wu, S. Zhao, J. Zhang, and C. Lu, "Remote sensing image sharpening by integrating multispectral image super-resolution and convolutional sparse representation fusion," *IEEE Access*, vol. 7, pp. 46562–46574, 2019.
[50] R. Fei, J. Zhang, J. Liu, F. Du, P. Chang, and J. Hu, "Convolutional sparse representation of injected details for pansharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 10, pp. 1595–1599, Oct. 2019.
[51] Q. Wang, Y. Zheng, G. Yang, W. Jin, X. Chen, and Y. Yin, "Multiscale rotation-invariant convolutional neural networks for lung texture classification," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 1, pp. 184–195, Jan. 2018.
[52] 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.
[53] V. Papyan, J. Sulam, and M. Elad, "Working locally thinking globally: Theoretical guarantees for convolutional sparse coding," *IEEE Trans. Signal Process.*, vol. 65, no. 21, pp. 5687–5701, Jan. 2017.
[54] F. Heide, W. Heidrich, and G. Wetzstein, "Fast and flexible convolutional sparse coding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 5135–5143.
[55] J.-F. Cai, E. J. Candes, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.
[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] X. Fu, Z. Lin, Y. Huang, and X. Ding, "A variational pan-sharpening with local gradient constraints," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 10265–10274.
[58] C. Jiang, H. Zhang, H. Shen, and L. Zhang, "Two-step sparse coding for the pan-sharpening of remote sensing images," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 7, no. 5, pp. 1792–1805, May 2014.
[59] 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.
[60] 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.
[61] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.
[62] 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.
[63] 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.

Kai Zhang was born in Shanxi, China, in 1992. He received the B.S. degree in electrical engineering and automation from 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 School, Shandong Normal University, Jinan, China. His research interests include multisource remote sensing image fusion, matrix factorization, and deep learning.

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 from the School of Information Science and Engineering School, Shandong Normal University.

Her research interests include deep learning and image processing.

Zhixi Feng received the B.S. degree in automation from Lanzhou University of Technology, Lanzhou, China, in 2012 and the Ph.D. degree in intelligent information processing from Xidian University, Xi'an, China, in 2018.

He is currently an Associate Professor with the School of Artificial Intelligence, Xidian University. His research interests include hyperspectral image processing and machine learning.

Jiande Sun received the Ph.D. degree in communication and information system from Shandong University, Jinan, 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. From 2014 to 2015, he was a DAAD Visiting Researcher with the Technical University of Berlin and the University of 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/coauthored more than 60 journal and conference papers. He is the coauthor of two books. His 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, Jinan, 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, the Executive Director of the Shandong Remote Sensing Society. He presided over or participated in three nation natural science foundations of China.