Fusion of Hyperspectral and Multispectral Images Accounting for Localized Inter-Image Changes

Xiyou Fu, *Member, IEEE*, Sen Jia, *Senior Member, IEEE*, Meng Xu, *Member, IEEE*,
Jun Zhou, *Senior Member, IEEE*, and Qingquan Li

Abstract—The high spectral resolution of hyperspectral images (HSIs) generally comes at the expense of low spatial resolution, which hinders the application of HSIs. Fusing an HSI and a multispectral image (MSI) from different sensors to get an image with the high spatial and spectral resolution is an economic and effective approach, but localized spatial and spectral changes between images acquired at different time instants can have negative impacts on the fusion results, which has rarely been considered in many fusion methods. In this article, we propose a novel group sparsity constrained fusion (GSFus) method to fuse hyperspectral and MSIs based on matrix factorization. Specifically, we imposed $\ell_{2,1}$ norm on the residual term of the MSI to account for the localized interimage changes occurring during the acquisition of the hyperspectral and MSIs. Furthermore, by exploiting the plug-and-play framework, we plugged a state-of-the-art denoiser, namely block-matching and 3-D filtering (BM3D), as the prior of the subspace coefficients. We refer to the proposed fusion method as GSFus method. We performed fusion experiments on two kinds of datasets, i.e., with and without obvious localized changes between the HSIs and MSIs, and a full resolution dataset. Extensive experiments in comparison with seven state-of-the-art fusion methods suggest that the proposed fusion method is more effective on fusing hyperspectral and MSIs than the competitors.

Index Terms—Group sparsity, hyperspectral image (HSI), image fusion, interimage changes, multispectral image (MSI).

I. Introduction

BY DIVIDING the electromagnetic wave into hundreds of spectral channels and measuring the electromagnetic energy scattered into the instantaneous field of view, hyperspectral cameras can produce hyperspectral images (HSIs)

with hundreds of spectral bands [1], [2]. The main advantage of HSIs over the traditional visible or multispectral images (MSIs) is the high spectral resolution, which is conducive to the precise recognition and identification of the materials present in the images and promotes remarkable applications in remote sensing, such as land cover classification [3], spectral unmixing [4], [5], change detection [6], and anomaly detection [7], [8].

The high spectral resolution of HSIs generally comes at the expense of low spatial resolution [9]. Due to the narrow bandwidth of the electromagnetic wave scattered into the instantaneous field of view, the spatial resolution of HSIs has to be lower so that more photons will be acquired. The low spatial resolution hinders the further application of HSIs [10]. In contrast, MSIs usually have a much higher spatial resolution and lower spectral resolution. In order to increase the spatial resolution of HSIs to obtain images with high spatial and spectral resolution at the same time, an economic and effective approach is to fuse a low-spatial-resolution HSI (LR-HSI) and a high-spatial-resolution MSI (HR-MSI) to produce a high-spatial-resolution HSI (HR-HSI) [11].

Perfect materials for fusion will be images of the same scene acquired at exactly the same conditions and at the same time instant. As the development of remote sensing, there emerge some platforms carrying both hyperspectral and multispectral imaging sensors, but the number of this kind of platforms is still limited [12]. On the contrary, optical satellites (such as Landsat, Sentinel, and Quickbird) can provide a large amount of high-resolution MSIs, which can be used to fuse with HSIs acquired by different sensors. The relatively short revisit cycles and the high spatial resolution of MSIs provide a great potential to fuse with HSIs for practical applications, such as the fusion of Landsat and moderate-resolution imaging spectroradiometer (MODIS) data to generate HSI sequences with a high temporal resolution [13], [14].

By fusing multiplatform hyperpsectral and MSIs, we can get images with high spatial and spectral resolution. However, multiplatform data fusion introduces further issues regarding data calibration and the consistency of the images to fuse [15]. In fact, crucial problems concern the different points of view and the different acquisition time that can be hardly compensated through the remotely sensed data-processing chain. But most of the research studies facing the problem of HSI fusion only include tests performed on simulated data under certain ideal assumptions, ignoring the described practical issues [12], [16], [17]. In particular, most of the fusion methods work under an assumption that the

observed images were acquired at almost the same observing conditions, implying that the LR-HSI and the HR-MSI can be viewed as spatially and spectrally downsampled versions of a unique underlying HR-HSI, respectively, [18]. However, this assumption is not always met in practical applications. It may happen that the observed HSI and MSI were acquired at long time intervals and different points of view. In this sense, the LR-HSI and the HR-MSI cannot be simply viewed as spatially and spectrally downsampled versions of a unique underlying HR-HSI because localized spatial and spectral changes will occur during the acquisition of the observed HSI and MSI. Spectral changes (termed as spectral variability in [19]) is an effect commonly observed in many scenes in which the spectral signatures of the pure constituent materials vary across the observed HSI. It can be caused under variable illumination and atmospheric conditions, and can also be intrinsic to the very definition of pure material, such as signatures of a single vegetation species varying significantly due to different growing and environmental conditions [19]–[21]. In addition, localized land-cover changes may occur during the time interval between the capture of the images. If not taken into consideration, the localized spatial and spectral changes between HSIs and MSIs can have a negative impact on the fusion results.

Considering the existence of localized spatial and spectral changes between the observed HSIs and MSIs, we propose a novel group sparsity constrained method to fuse hyperspectral and MSIs based on matrix factorization. In order to accounting for localized changes between the hyperspectral and multispectra images, we characterize the spectral difference between the observed LR-MSI and the underlying HR-HSI by a group sparsity norm, $\ell_{2,1}$ norm. To exploit the intrinsic spatial information in the subspace coefficients and avoid cumbersomely designing a complex prior for subspace coefficients, we propose the use of the plug-and-play technique, i.e., using a plugged prior for the subspace coefficients. We plugged the image prior of block-matching and 3-D filtering (BM3D), which is a state-of-the-art denoiser, in the model to regularize the subspace coefficients. We refer to the proposed hyperspectral and MSIs fusion method as group sparsity constrained fusion (GSFus) method. The main contributions of this article are summarized as follows.

1) To fuse hyperspectral and MSIs, we proposed a novel GSFus method, which can address localized changes between the observed images by using a form of group sparsity. Specifically, by imposing the $\ell_{2,1}$ norm on the residual term of the MSI to promote the group sparsity, the proposed fusion method can account for the localized changes occurring between the hyperspectral and MSIs obtained from different sensors at long time intervals.

2) The spectral low-rankness and spatial self-similarity of the underlying HR-HSI are exploited by using a subspace representation and plug-and-play technique, respectively. To avoid cumbersomely designing a hand-crafted regularizer for the subspace coefficients of the underlying HR-HSI, we plugged the image prior from BM3D, which is a state-of-the-art denoiser, in the model.

The plugged prior was proven to be effective in regularizing the subspace coefficients.

3) We performed fusion experiments on two kinds of datasets, i.e., with and without obvious localized changes between the observed HSIs and MSIs, and a full resolution dataset. Extensive experiments in comparison with seven state-of-the-art fusion methods suggest that the proposed fusion method is more effective on fusing hyperspectral and MSIs.

We organized the remainder of this article as follows. We first briefly introduce the related work of hyperspectral and MSIs fusion in Section II. Then, we introduce our proposed fusion method in detail in Section III. In Section IV, we show and analyze the experimental results of the proposed method and the competitors. Finally, we make a conclusion of this article in Section V.

II. RELATED WORK

Fusion of an HSI and an MSI to reconstruct a high-spatial-resolution HSI is an economic way and has attracted more and more interest. The existing HSI and MSI fusion methods can be approximately classified into matrix factorization-based methods, tensor factorization-based methods, deep learning-based methods, and pan-sharpening methods extended for fusion of HSIs and MSIs.

Matrix factorization-based methods fuse the images by decomposing the target HR-HSI into several matrices. Based on the assumption that each spectral pixel can be sparsely represented by the spectral basis and the corresponding coefficients. Sparse representation-based methods fuse the images by exploiting the sparsity of the coefficients. Kawakami et al. [22] fused the images by estimating the spectral basis from the LR-HSI and conducting sparse coding for HR-MSI. Dong et al. [23] estimated the coefficients using a structured sparse coding method. The unmixing-based fusion methods first extract the endmembers as the spectral basis, and then alternately update the endmembers via LR-HSI and coefficients via HR-MSI by introducing different unmixing priors to regularize the coefficients. Yokoya et al. [24] proposed a coupled nonnegative matrix factorization (CNMF) unmixing method, in which hyperspectral and multispectral data are alternately decomposed into endmember and abundance matrices based on a linear spectral mixture model. Bendoumi et al. [25] proposed a fusion framework in which the whole image is divided into several subimages and the spectral-unmixing-based fusion algorithm is executed on the subimages. Lanaras et al. [26] proposed a fusion method by jointly unmixing two input images into pure reflectance spectra and corresponding coefficients with a number of physical constraints. The subspace representation models obtain spectral basis by performing singular value decomposition (SVD) or vertex component analysis (VCA) on the LR-HSI, and then estimate the coefficients by imposing different priors on the coefficients, such as maximum a posteriori (MAP) [27], vector total variation regularization [28], sparse representation-based regularizer [29], and low tensor multirank (LTMR) regularization [30]. Recently, Borsoi et al. [18] proposed a matrix
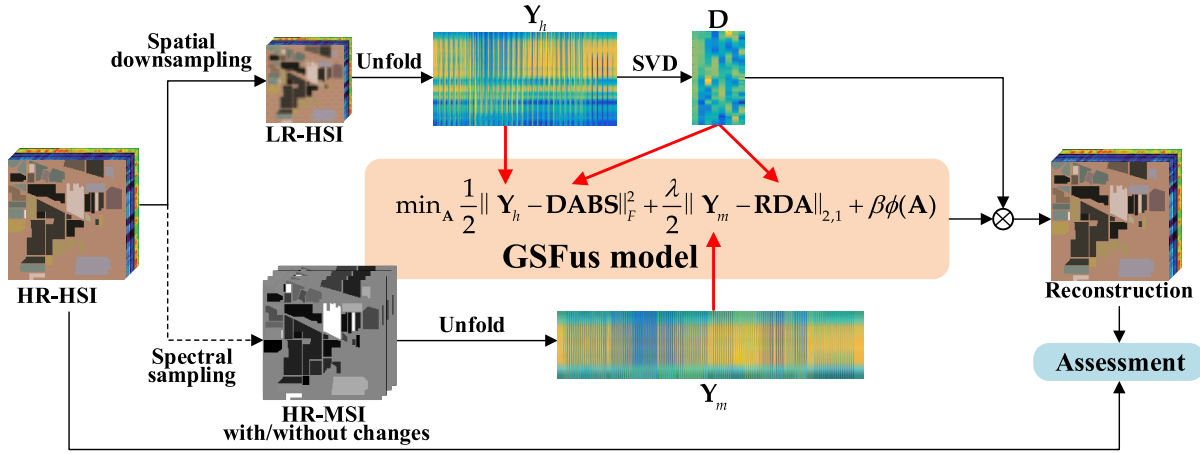
Fig. 1. Flowchart of the proposed fusion method.

factorization-based fusion method accounting for spectral variability, leading to performance improvements in the presence of spectral variability.

Tensor factorization-based methods are a kind of methods that treat the images as a tensor to preserve the spatial and spectral structure of the images rather than reshaping them into matrices. In order to preserve the spatial and spectral smoothness, Zhang *et al.* [31] proposed a low-rank Tucker decomposition model in a combination of two graphs to fuse HSI and MSI. By taking the HR-HSI as a 3-D tensor, Li *et al.* [32] proposed solving the fusion problem by estimating the core tensor and the dictionaries along three modes. Kanatsoulis *et al.* [33] proposed a coupled tensor factorization framework that guarantees the identifiability of the super-resolution image under mild and realistic conditions. Xu *et al.* [34] proposed a nonlocal tensor-based fusion model based on t-SVD, which used a different tensor sparse representation model from [32]. By grouping the image into 4-D tensors and introducing the tensor-train rank to regularize the 4-D tensors, Dian *et al.* [35] proposed the low tensor-train rank representation fusion method. Chang *et al.* [36] proposed an image restoration method based on weighted low-rank tensor decomposition. In this work, regularization parameters on the core tensor were given based on the different singular values. Bu *et al.* [37] proposed a novel graph Laplacian-guided coupled tensor decomposition model for fusion of HSIs and MSIs. He *et al.* [38] proposed a coupled tensor ring factorization fusion model, which simultaneously learns the tensor ring core tensors of the HR-HSI from a pair of HSI and MSI. Borsoi *et al.* [39] proposed a coupled tensor approximation fusion method accounting for both spatially and spectrally localized changes.

Due to the powerful nonlinear learning capability, deep learning has attracted great attention and has been successfully utilized in many image processing applications. In terms of deep-learning-based methods for fusion of HR-MSI and LR-HSI, Palsson *et al.* [40] first proposed a 3-D convolutional neural network (CNN) method to fuse HSIs and MSIs, by learning the spatial decimation filter. Dian *et al.* [41] incorporated the deep learning prior into the fusion framework and fused the images by solving a Sylvester equation. Zhou *et al.* [42] proposed a pyramid fully CNN-based method, which is composed of an encoder and a pyramid fusion subnetworks. By formulating the fusion of HSI with MSI as a nonlinear spectral mapping from an HSI to MSI and clustering the spectral bands into several groups, Han *et al.* [43] proposed the multibranch back propagation (BP) neural network-based fusion method. To fully use the information of HSIs and MSIs, Yang *et al.* [44] proposed a novel two branches CNN method, which extracts the features from the spectrum of each pixel in low-resolution HSI, and its corresponding spatial neighborhood in MSIs, respectively. Xue *et al.* [45] proposed a variational fusion network, in which the degradation model and data prior are implicitly represented by a deep learning network and jointly learned from the training data.

Over the past two decades, a large number of pan-sharpening techniques have been developed to enhance the spatial resolution of MSIs, including component substitution -based methods [46], [47], multiresolution analysis-based methods [48], [49], and sparse representation-based methods [50], [51]. As pan sharpening can be considered a special case of the hyperspectral and multispectral fusion problem, efforts have also been made toward generalizing existing pan-sharpening methods for fusion of HSIs and MSIs. One of the first attempts of pan-sharpening-based hyperspectral and multispectral fusion was presented in [52] using a wavelet technique. A successful solution consists in assigning to each hyperspectral band, whose enhancement is separately performed, a single channel of the multispectral data [53], [54]. Instead of choosing a band, Chen *et al.* [55] proposed a framework that solves the fusion problem by dividing the spectrum of hyperspectral data into several regions and fusing hyperspectral and multispectral images in each region using conventional pan-sharpening techniques. Selva *et al.* [56] proposed a framework, called *hypersharpening*, that utilizes a weighted combination of all the multispectral bands for the spatial improvement of each hyperspectral

band, achieving significantly better fusion results than simply selecting a band from MSIs.

III. PROPOSED GSFUS METHOD

From Sections I and II, we can see that it is important to incorporate the localized changes between the observed HSIs and MSIs when we fuse the images from different sensors. In addition, it is important to define a prior that can fully make use of the spatial information embedded in the subspace coefficients. Thus, in this article, we propose a GSFus method to fuse the HSIs and MSIs based on matrix factorization. In this method, we promoted the group sparsity of the reconstruction error of the MSI term by minimizing its $\ell_{2,1}$ norm. Furthermore, by introducing the plug-and-play framework, we plugged a state-of-the-art denoiser, i.e., BM3D, as the prior to regularize subspace coefficients. The flowchart of the proposed fusion method is illustrated in Fig. 1.

A. Observation Model

Let $\mathcal{Y}_h \in \mathbb{R}^{w \times h \times L}$ denote the LR-HSI observed by a hyperspectral imaging sensor, and $\mathcal{Y}_m \in \mathbb{R}^{W \times H \times l}$ denote the HR-MSI observed by a multispectral imaging sensor, where w, h denote the number of rows, columns in LR-HSI, and l denote the number bands in HR-MSI, respectively. Let $\mathcal{Z} \in \mathbb{R}^{W \times H \times L}$ denote the underlying HR-HSI of the observed LR-HSI scene, where W, H, L denote the number of rows, columns and bands in the HR-HSI, respectively. We have $W > w$, $H > h$, and $L > l$. The three-dimensional HSI image can be reshaped into a matrix by unfolding along the spectral mode. Here, we use $\mathbf{Z} \in \mathbb{R}^{L \times WH}$, $\mathbf{Y}_h \in \mathbb{R}^{L \times wh}$, $\mathbf{Y}_m \in \mathbb{R}^{l \times WH}$ to denote the unfolding of the HR-HSI, LR-HSI, and HR-MSI, respectively.

As we aim to fuse an HSI and an MSI of the same scene from different sensors at different time instants, the LR-HSI $\mathbf{Y}_h$ acquired at its acquisition time can be represented as the spatially downsampled version of HR-HSI, $\mathbf{Z}$, as

$$\mathbf{Y}_h = \mathbf{ZBS} + \mathbf{N}_h \tag{1}$$

where $\mathbf{B} \in \mathbb{R}^{WH \times WH}$ is a blur kernel that encapsulates the hyperspectral sensor's spatial point-spread function (PSF) describing the sensor's spatial degradation process, and $\mathbf{S} \in \mathbb{R}^{WH \times wh}$ is the downsampling matrix, and $\mathbf{N}_h \in \mathbb{R}^{L \times wh}$ is additive Gaussian noise. The blur kernel $\mathbf{B}$ can be decomposed as

$$\mathbf{B} = \mathbf{FKF}^{-1} \tag{2}$$

where $\mathbf{F}$ and $\mathbf{F}^{-1}$ denote the fast Fourier transform (FFT) and the inverse FFT, respectively. The matrix $\mathbf{K}$ is a diagonal matrix whose diagonal values are the eigenvalues of $\mathbf{B}$.

The HR-MSI, $\mathbf{Y}_m$, is the spectral downsampled version of $\mathbf{Z}$ and can be represented as

$$\mathbf{Y}_m = \mathbf{RZ} + \mathbf{N}_m \tag{3}$$

where $\mathbf{R} \in \mathbb{R}^{l \times L}$ is the spectral response function, and $\mathbf{N}_m \in \mathbb{R}^{l \times WH}$. If there is no spectral variability [18] or interimage changes [39] occurring between the acquisition of the observed LR-HSI and HR-MSI, $\mathbf{N}_m$ represents additive Gaussian noise. However, if there exists localized spatial and spectral changes
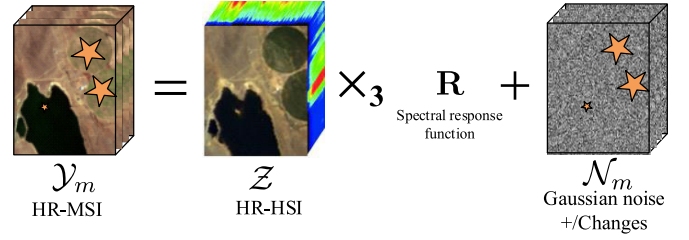

Fig. 2. Tensor representation of the observation model (3). The brown stars mark the areas of changes, and $\mathcal{N}_m$ is the tensor form of $\mathbf{N}_m$. $\times_3$ denotes the mode-n product of a tensor by a matrix.
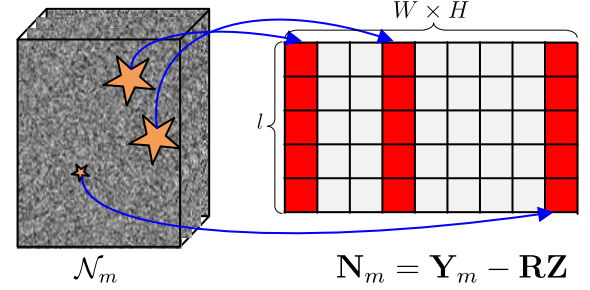

Fig. 3. Illustration of group sparsity of $\mathbf{N}_m$ when image changes occur during the acquisition of the LR-HSI and HR-MSI. The brown stars mark the areas of changes, and $\mathcal{N}_m$ is the tensor form of $\mathbf{N}_m$. The gray squares represent the elements close to zero and the red squares represent the nonzero elements.

occurring between the acquisition of the observed LR-HSI and HR-MSI, $\mathbf{N}_m$ accounts for not only additive Gaussian noise but also interimage changes. The observation model (3) is depicted in Fig. 2. In this scenario, $\mathbf{N}_m$ is a matrix with group sparsity characteristics, as illustrated in Fig. 3.

B. Localized Spatial and Spectral Changes Between Observed Images

The localized spatial and spectral changes (called interimage changes in [39]) between observed images are usually caused by large acquisition time differences. It happens that sometimes we have to fuse two images acquired at significantly different times. Note that land-cover of Earth surface is changing over time. This motivates us to develop more flexible fusion models.

To illustrate the existence of localized spatial and spectral changes between observed images, we use three pairs of real hyperspectral and real MSIs, namely Paris data, Lake Tahoe data, and Ivanpah Playa data, as examples (shown in Fig. 4). A brief introduction of these data can be found in Section IV. To see the interimage changes in each pair images, we generated a difference image, $\mathbf{N}_m$, by computing $\mathbf{RZ} - \mathbf{Y}_m$, i.e., downsampling the HSI, $\mathbf{Z}$, and then subtracting it from the MSI, $\mathbf{Y}_m$. If no obvious interimage changes exist, then $\mathbf{N}_m$ accounts for only Gaussian noise (which is pixel-wise and band-wise independent). In this scenario, elements of a matrix $\mathbf{1}_l^T \mathbf{N}_m$ are also Gaussian distributed. Histogram plots of $\mathbf{1}_l^T \mathbf{N}_m$ for their images are given in the third column of Fig. 4, where we can see that the histogram for Paris data is approximately Gaussian, implying no obvious interimage
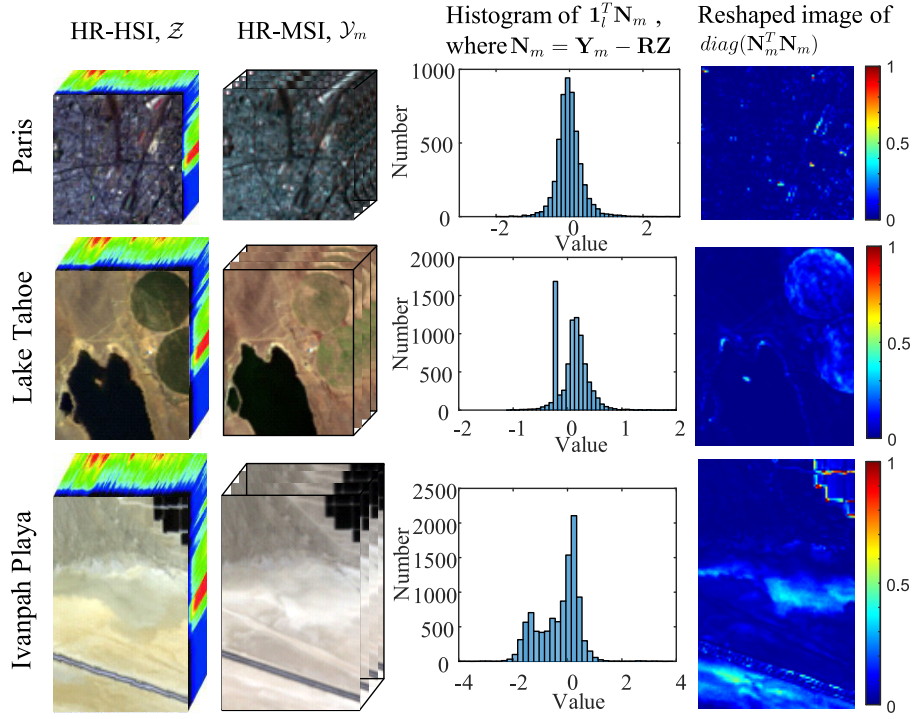
Fig. 4. Illustration of the existence of localized spatial and spectral changes between observed images.

changes exist. On the other hand, the histograms for Lake Tahoe and Ivanpah Playa data are non-Gaussian, meaning that significant interimage changes exist in these two pairs. The interimage changes may also be inferred in the final column of Fig. 4, where displays power or energy of each pixel in the difference image, $\mathbf{N}_m$. We obtained signal power by computing $\text{diag}(\mathbf{N}_m^T\mathbf{N}_m)$,[1] followed by reshaping the column vector into a $W \times H$ matrix. As we can see from the final column of Fig. 4, pixels of large values indicate significant changes between HR-HSI and HR-MSI. The Paris scene exhibits very minor changes because its images were acquired at the same time instant [28]. The Lake Tahoe scene and Ivanpah Playa scene show significant and localized changes, which is mainly due to large acquisition time differences. The pair of Lake Tahoe scenes were acquired on October 4, 2014, and October 24, 2017, and on October 26, 2015, and December 17, 2017, respectively.

C. Estimation of Subspace

HSIs have a high correlation in the spectral domain, which means that the spectral vectors live in low-dimensional manifolds or subspaces. As a result, the HR-HSI can be represented by the subspace and its corresponding coefficients as shown in Fig. 5, which can be written as

$$\mathbf{Z} = \mathbf{D}\mathbf{A} \quad (4)$$

where $\mathbf{D} \in \mathbb{R}^{L \times k}$, $\mathbf{A} \in \mathbb{R}^{k \times WH}$ are the subspace and the corresponding coefficients, respectively.

[1]Function $\text{diag}(\mathbf{X})$ outputs a column vector of the main diagonal elements of $\mathbf{X}$.
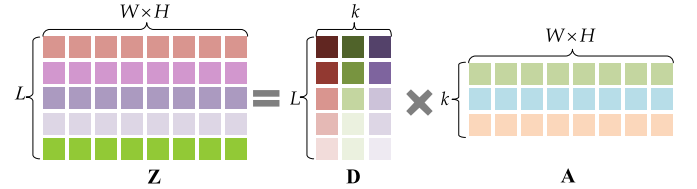
Fig. 5. Subspace representation of the HR-HSI $\mathbf{Z}$.

As LR-HSI can be regarded as the spatial downsampling version of the HR-HSI, we can assume that the spectral information that remained in the LR-HSI and HR-HSI is the same, which means that the LR-HSI and HR-HSI share the same subspace. Thus, we can estimate the subspace from LR-HSI using subspace learning methods, i.e., SVD. The SVD of LR-HSI can be expressed as

$$\mathbf{Y}_h = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T \quad (5)$$

where $\mathbf{U} \in \mathbb{R}^{L \times L}$ and $\mathbf{V} \in \mathbb{R}^{WH \times WH}$ are semiunitary matrices, and $\boldsymbol{\Sigma} \in \mathbb{R}^{L \times WH}$ are diagonal matrix which contains singular values ranking by nonincreasing order. By discarding $(L - k)$ smallest singular values and keeping the k largest ones of $\boldsymbol{\Sigma}$, we can obtain $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(:, 1 : k)$. Then, a low-rank approximation of $\mathbf{Y}_h$ can be obtained as follows:

$$\hat{\mathbf{Y}}_h = \hat{\mathbf{U}}\hat{\boldsymbol{\Sigma}}\hat{\mathbf{V}}^T \quad (6)$$

where $\hat{\mathbf{U}} = \mathbf{U}(:, 1 : k)$, $\hat{\mathbf{V}} = \mathbf{U}(:, 1 : k)$. The subspace can be obtained as

$$\mathbf{D} = \hat{\mathbf{U}} = \mathbf{U}(:, 1 : k). \quad (7)$$

D. GSFus Model

The objective function of the proposed GSFus method is written as

$$\min_{\mathbf{A}} \frac{1}{2}\|\mathbf{Y}_h - \mathbf{DABS}\|_F^2 + \frac{\lambda}{2}\|\mathbf{Y}_m - \mathbf{RDA}\|_{2,1} + \beta\phi(\mathbf{A}) \quad (8)$$

where $\phi(\cdot)$ is the regularizer defined to further enhance the spatial correlation of subspace coefficients $\mathbf{A}$, $\lambda > 0$ and $\beta > 0$ are parameters to regularize the group sparsity property of $(\mathbf{Y}_m - \mathbf{RDA})$ and the image prior of $\mathbf{A}$. The mixed norm $\ell_{2,1}$ norm of $\mathbf{X}$ is defined as $\|\mathbf{X}\|_{2,1} = \sum_{i=1}^{n}\|\mathbf{x}_i\|_2$ ($\mathbf{x}_i$ represents ith column vector of $\mathbf{X}$). The mixed norm promotes column-wise sparsity of $(\mathbf{Y}_m - \mathbf{RDA})$.

Note that in (8), we introduce the $\ell_{2,1}$ norm on the residual term of the MSI, which is quite different from existing fusion methods [28], [30]. The use of $\ell_{2,1}$ norm, instead of Frobenius norm, promotes group sparsity and enables us to characterize the spatial and spectral changes between observed MSI, $\mathbf{Y}_m$, and the underlying HR-HSI, $\mathbf{DA}$. Because of the $\ell_{2,1}$ norm, the minimizer of (8) involves the promotion of group sparsity. In this sense, the estimation of $\mathbf{A}$ at the positions of no changes will be fitted using the information from both the LR-HSI and HR-MSI. Whereas, the estimation of $\mathbf{A}$ at the positions that changes take place will be determined by the fidelity term of the LR-HSI and the regularization term on $\mathbf{A}$, which will help to achieve good reconstruction results. The $\ell_{2,1}$ norm has been successfully used in image destripping methods [57], [58] to promote group sparsity of matrices, representing image stripes, and also used to promote group sparsity of anomaly matrices in anomaly detection tasks [7], [8]. To the best of our knowledge, this is the first time that $\ell_{2,1}$ norm is used to address the hyperspectral and MSIs fusion problems.

In addition, by exploiting the plug-and-play techniques, we can use the a plugged prior as the implicit regularizer on $\mathbf{A}$, which means that we do not need to define the function $\phi(\cdot)$ explicitly. The plugged priors have been successfully used in hyperspectral inverse problems to produce outstanding results, e.g., compressive sensing [59], anomaly detection [7], [8], image denoising [60], [61] and so on. Here, we used state-of-the-art denoisers, BM3D [62], to regularize $\mathbf{A}$. By simultaneously making use of the sparsity and the nonlocal self-similarity of images, BM3D can achieve satisfying denoising performance while preserving the image details. The good generalization ability and efficient implementation also make it a preferred denoiser to be plugged.

We introduce two auxiliary variables, i.e., $\mathbf{V}_1$ and $\mathbf{V}_2$, and convert (8) into a constrained optimization as

$$\min_{\mathbf{A},\mathbf{V}_1,\mathbf{V}_2} \frac{1}{2}\|\mathbf{Y}_h - \mathbf{DABS}\|_F^2 + \frac{\lambda}{2}\|\mathbf{V}_1\|_{2,1} + \beta\phi(\mathbf{V}_2)$$
$$\text{s.t. } \mathbf{V}_1 = \mathbf{Y}_m - \mathbf{RDA}, \quad \mathbf{V}_2 = \mathbf{A}. \quad (9)$$

Then, we can easily transform the constraint problem into an optimization of the following augmented

Lagrangian function:

$$L(\mathbf{A}, \mathbf{V}_1, \mathbf{V}_2, \mathbf{G}_1, \mathbf{G}_2)$$
$$= \frac{1}{2}\|\mathbf{Y}_h - \mathbf{DABS}\|_F^2 + \frac{\lambda}{2}\|\mathbf{V}_1\|_{2,1}$$
$$+ \beta\phi(\mathbf{V}_2) + \frac{\mu}{2}\left\|\mathbf{V}_1 - \mathbf{Y}_m + \mathbf{RDA} + \frac{\mathbf{G}_1}{\mu}\right\|_F^2$$
$$+ \frac{\mu}{2}\left\|\mathbf{V}_2 - \mathbf{A} + \frac{\mathbf{G}_2}{\mu}\right\|_F^2 \quad (10)$$

where μ denotes the penalty parameter and $\mathbf{G}_1$, $\mathbf{G}_2$ are the Lagrangian Multipliers. By referring to the alternative method of multipliers (ADMMs), we can split (10) into several subproblems and solve each subproblem alternatively.

1) Subproblem of $\mathbf{A}$: Subproblem of $\mathbf{A}$ can be solved by minimizing $L(\mathbf{A}, \mathbf{V}_1, \mathbf{V}_2, \mathbf{G}_1, \mathbf{G}_2)$ w.r.t. $\mathbf{A}$ as

$$\mathbf{A} \in \arg\min_{\mathbf{A}} L(\mathbf{A}, \mathbf{V}_1, \mathbf{V}_2, \mathbf{G}_1, \mathbf{G}_2)$$
$$= \arg\min_{\mathbf{A}} \frac{1}{2}\|\mathbf{Y}_h - \mathbf{DABS}\|_F^2$$
$$+ \frac{\mu}{2}\left\|\mathbf{V}_1 - \mathbf{Y}_m + \mathbf{RDA} + \frac{\mathbf{G}_1}{\mu}\right\|_F^2$$
$$+ \frac{\mu}{2}\left\|\mathbf{V}_2 - \mathbf{A} + \frac{\mathbf{G}_2}{\mu}\right\|_F^2. \quad (11)$$

Since (11) is strongly convex, the solver of $\mathbf{A}$ can be obtained by setting the derivative of (11) w.r.t. $\mathbf{A}$ to zero. Since columns of $\mathbf{D}$ are orthogonal, it satisfies $\mathbf{D}^T\mathbf{D} = \mathbf{I}$. We can finally acquire a Sylvester equation [63]

$$\mathbf{H}_1\mathbf{A} + \mathbf{A}\mathbf{H}_2 = \mathbf{H}_3 \quad (12)$$

where

$$\mathbf{H}_1 = \mu(\mathbf{RD})^T\mathbf{RD} + \mu\mathbf{I}$$
$$\mathbf{H}_2 = (\mathbf{BS})(\mathbf{BS})^T$$
$$\mathbf{H}_3 = \mathbf{D}^T\mathbf{Y}_h(\mathbf{BS})^T + \mu\left(\mathbf{V}_2 + \frac{\mathbf{G}_2}{\mu}\right)$$
$$- \mu(\mathbf{RD})^T\left(\mathbf{V}_1 - \mathbf{Y}_m + \frac{\mathbf{G}_1}{\mu}\right). \quad (13)$$

Although (12) can be solved by referring to the conjugate gradient method, in this article, we solved (12) analytically and efficiently as done in [11] and [30].

2) Subproblem of $\mathbf{V}_1$: Subproblem of $\mathbf{V}_1$ can be solved by minimizing $L(\mathbf{A}, \mathbf{V}_1, \mathbf{V}_2, \mathbf{G}_1, \mathbf{G}_2)$ w.r.t. $\mathbf{V}_1$ as

$$\mathbf{V}_1 \in \arg\min_{\mathbf{V}_1} L(\mathbf{A}, \mathbf{V}_1, \mathbf{V}_2, \mathbf{G}_1, \mathbf{G}_2)$$
$$= \arg\min_{\mathbf{V}_1} \frac{\lambda}{2}\|\mathbf{V}_1\|_{2,1} + \frac{\mu}{2}\left\|\mathbf{V}_1 - \mathbf{Y}_m + \mathbf{RDA} + \frac{\mathbf{G}_1}{\mu}\right\|_F^2 \quad (14)$$

which has a close-form solution by using the $\ell_{2,1}$ norm minimization operator [64]

$$\mathbf{V}_1 = \text{shrink}_{2,1}\left(\mathbf{Y}_m - \mathbf{RDA} - \frac{\mathbf{G}_1}{\mu}, \frac{\lambda}{2\mu}\right) \quad (15)$$

where

$$\text{shrink}_{2,1}(\mathbf{U}, \tau) = [\text{shrink}(\mathbf{u}_i, \tau), i = 1, \ldots, N]. \quad (16)$$
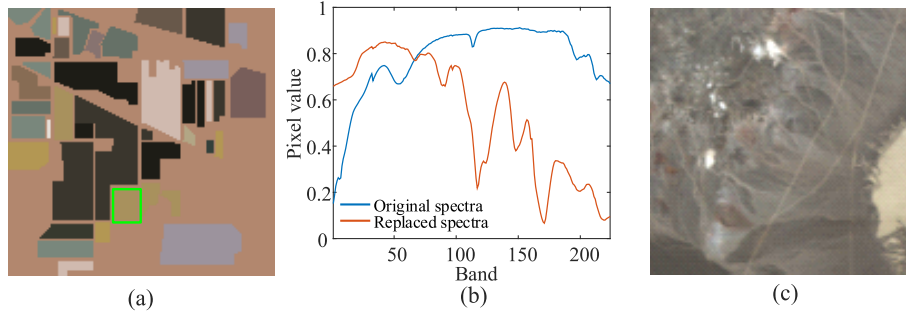
Fig. 6. Simulated Indian Pines data and Cuprite data used in the experiments. (a) Simulated Indian Pines data (Composed of bands 28, 19, and 10). (b) Original spectra and replaced spectra of the selected region as marked in the green box in (a). (c) Cuprite data (Composed of bands 32, 19, and 10).

Here, shrink($\mathbf{u}, \tau$) is used to denote a vector-soft-threshold operation and can be expressed as

$$\mathbf{u} \mapsto \frac{\max(\|\mathbf{u}\| - \tau, \mathbf{0})}{\max(\|\mathbf{u}\| - \tau, \mathbf{0}) + \tau} \mathbf{u}.$$

3) Subproblem of $\mathbf{V}_2$: Subproblem of $\mathbf{V}_2$ can be solved by minimizing $L(\mathbf{A}, \mathbf{V}_1, \mathbf{V}_2, \mathbf{G}_1, \mathbf{G}_2)$ w.r.t. $\mathbf{V}_2$ as

$$\mathbf{V}_2 \in \arg\min_{\mathbf{V}_2} L(\mathbf{A}, \mathbf{V}_1, \mathbf{V}_2, \mathbf{G}_1, \mathbf{G}_2)$$

$$= \arg\min_{\mathbf{V}_2} \beta \phi(\mathbf{V}_2) + \frac{\mu}{2} \left\| \mathbf{V}_2 - \mathbf{A} + \frac{\mathbf{G}_2}{\mu} \right\|_F^2. \quad (17)$$

Based on the plug-and-play technique, the solution of the problem (17) can be obtained by using denoising operator [65] as

$$\mathbf{V}_2 \leftarrow \text{Denoiser} \left(\mathbf{A} - \frac{\mathbf{G}_2}{\mu}, \frac{\beta}{\mu} \right) \quad (18)$$

where Denoiser($\cdot$) is a plugged denoiser. By exploiting plug-and-play technique, we can plug in an off-the-shelf denoiser, such as BM3D [62], weighted nuclear norm minimization (WNNM) [66], and CNN-based denoising methods [67]), rather than making effort to design an effective regularizer. In this article, we use the BM3D as the plugged denoiser to solve the problem (18), as it is a fast and state-of-the-art denoiser.

4) Updating of $\mathbf{G}_1$: We update $\mathbf{G}_1$ in each iteration as

$$\mathbf{G}_1 = \mathbf{G}_1 + \mu(\mathbf{V}_1 - \mathbf{Y}_m + \mathbf{RDA}). \quad (19)$$

5) Updating of $\mathbf{G}_2$: We update $\mathbf{G}_2$ in each iteration as

$$\mathbf{G}_2 = \mathbf{G}_2 + \mu(\mathbf{V}_2 - \mathbf{A}). \quad (20)$$

For better understanding, we summarized the steps for optimizing the objective function (8) in Algorithm 1.

E. Computational Complexity

The computational complexity of the proposed algorithm is given as follows. The computational complexity of estimating the subspace $\mathbf{D}$ is $\mathcal{O}(w^2 h^2 L)$. For each iteration, the computational complexity of updating $\mathbf{A}$ via (12) is $\mathcal{O}(WH \cdot \max\{\log(WH), k^2\})$ [63]. Updating $\mathbf{V}_1$ and $\mathbf{G}_1$ costs $\mathcal{O}(lkWH)$. The computational complexity of updating $\mathbf{V}_2$ is $\mathcal{O}(kWH + k\delta)$, where δ represents the computational complexity of denoising an eigenimage. Updating $\mathbf{G}_2$ costs $\mathcal{O}(WHk)$. The calculation of $\mathbf{Z}$ costs $\mathcal{O}(LkWH)$. Thus, the computational

---

Algorithm 1 Proposed GSFus Model

1: **Input:** $\mathbf{Y}_h, \mathbf{Y}_m, \lambda, \beta, k$
2: Estimate the subspace $\mathbf{D}$ via (7)
3: **While** not converge **do**
4:     Update $\mathbf{A}$ via (12)
5:     Update $\mathbf{V}_1$ via (15)
6:     Update $\mathbf{V}_2$ via (18)
7:     Update $\mathbf{G}_1$ via (19)
8:     Update $\mathbf{G}_2$ via (20)
9: **end while**
10: $\mathbf{Z} = \mathbf{DA}$
11: **Output $\mathbf{Z}$**

---

complexity of the whole proposed method is $\mathcal{O}(w^2 h^2 L + p(WH \cdot \max\{\log(WH), k^2\} + k\delta + \text{LkWH}))$, where p is the number of iterations. We performed experiments in MATLAB (R2016a) on Windows 10 with an Intel Core i7-7700HQ 2.80-GHz processor and 24-GB RAM. The running times for the experiments are presented in Tables III and IV.

IV. EXPERIMENTS

In this section, we present extensive experimental results to assess the performance of the proposed method for fusing HSIs and MSIs in comparison with seven state-of-the-art fusion methods. The experimental results were assessed qualitatively and quantitatively.

A. Experimental Datasets

We evaluated the fusion performance of the proposed fusion method using five hyperspectral datasets. In order to fully evaluate the proposed method, experiments were performed on datasets without obvious changes between HSIs and MSIs, as well as datasets that exhibit localized changes between HSIs and MSIs. We give the detail of the datasets used as follows.

1) The first dataset is the simulated Indian Pines dataset, which was synthesized using the spatial structure of the real Indian Pines image and spectral signatures extracted from the United States Geological Survey (USGS) digital spectral library [68]. The HR-HSI image of size 144 (rows) × 144 (columns) × 224 (bands) is shown in Fig. 6(a). We followed the Wald's protocol [69] to simulate the images. Specifically, the LR-HSI was

generated by applying a Gaussian filter to the HR-HSI, followed by a downsampling step with downsampling ratio equal to 2. The Gaussian filter was designed with its full-width at half-maximum (FWHM) equal to the downsampling ratio, to match a plausible system modulation transfer function (MTF) [12], [70], [71]. White Gaussian noise WGN) was added to the LR-HSI to generate a noisy observed LR-HSI, yielding a signal-to-noise ratio (SNR) equal to 30 dB. An HR-MSI with ten bands was generated by computing **RZ**, where **R** is a spectral response function of the Sentinel-2A instrument. To simulate a noisy observed HR-MSI commonly seen in real cases, WGN was added to the reference MSI to obtain an observed HR-MSI with SNR equal to 40 dB [12], [28], [39]. Furthermore, we also generate an observed HR-MSI with interimage changes as follows.

2) We selected a region of the HR-HSI as marked by the green box in Fig. 6(a), and then replaced the original spectra within the region with new spectra as shown in Fig. 6(b).

3) The observed HR-MSI with ten bands was generated by spectral downsampling the HR-HSI with replaced spectra using the spectral response of the Sentinel-2A instrument and then adding the WGN to generate a 40-dB SNR. Now, the Indian pine data has been used to simulate a pair of images with interimage changes and a pair of images without changes.

4) The second dataset is the Cuprite data captured by airborne visible infrared imaging spectrometer (AVIRIS) in Nevada with a spatial resolution of 30 m [72]. The image used for experiments has a size of 120×120 pixels and 196 bands. The HSI is shown in Fig. 6(c). The LR-HSI was generated as the first dataset. WGN was added to generate a noisy observed LR-HSI with a 30-dB SNR. We generated the HR-MSI by implementing spectral degradation of HR-HSI, i.e., **RZ**, where **R** is the spectral response of the Sentinel-2A instrument. WGN was added to a noisy observed HR-MSI, yielding a 40-dB SNR.

5) The third dataset is a pair of real hyperspectral and real MSIs acquired above Paris with a spatial resolution of 30 m [28]. The HSI was captured by the Hyperion and the MSI was obtained by the Advanced Land Imager instruments on board the Earth Observing-1 Mission satellite. The hyperspectral and MSIs are shown in Fig. 4. The reference HSI has a size of 72×72 pixels and 128 spectral bands. The MSI has nine bands. The LR-HSI was generated as the first dataset using the reference HSI. Finally, WGN was added to generate an observed LR-HSI with a 30-dB SNR. The observed HR-MSI was obtained by adding WGN to the reference MSI to obtain a 40-dB SNR. The spectral response function was estimated as described in [18] and [28].

6) The fourth dataset is the Lake Tahoe data, which is composed of an HSI and an MSI [18]. The HSI was acquired by the AVIRIS instrument on October 4, 2014, with a spatial resolution of 20 m. The MSI was captured by the Sentinel-2A instrument on October 24, 2017 with

a spatial resolution of 20 m. The HSI has a size of 100×80 pixels and contains 173 bands after preprocessing while the MSI contains ten bands. We show the HSI and MSI in Fig. 4. As acquired with a long time interval, significant changes can be seen from the two images. The spectral response function was obtained from calibration measurements as a known prior [18]. The observed LR-HSI and the observed HR-MSI were generated as the third dataset.

7) The fifth dataset is the Ivanpah Playa data, which is composed of a hyperspectral image, acquired by the AVIRIS instrument on October 26, 2015, and an MSI, captured by the Sentinel-2A instrument on December 17, 2017 [18]. Both images have a spatial resolution of 20 m with a size of 128×80. The HSI contains 173 bands after preprocessing while the MSI contains ten bands. The HSI and MSI are shown in Fig. 4. The spectral response function was obtained from calibration measurements as a known prior [18]. As acquired with a long time interval, significant changes can be seen from the two images. The observed LR-HSI, observed HR-MSI were generated as the third dataset.

8) The sixth dataset is a pair of hyperspectral and MSIs at full resolution. The LR-HSI was captured by the Hyperion sensor onboard the Earth Observing-1 satellite and has a spatial resolution of 30 m and 220 spectral bands within the spectral range of 400–2500 nm. After removing the bands of low SNR, 87 bands were preserved. We used an area of 100×100 pixels for fusion in this experiment. The HR-MSI was obtained by the Sentinel-2A satellite and has 13 spectral bands. Four bands of 10-m spatial resolution were kept for the fusion. The spatial size of the HR-MSI is 300×300.

B. Compared Methods and Evaluation Indexes

1) Compared Methods: To evaluate the performance of the proposed method, we compared the proposed method with seven state-of-the-art fusion methods, including matrix factorization-based methods, tensor factorization-based methods, and an adapted hyperspectral pansharpening method. Generalized Laplacian pyramid hypersharpening (GLPHS) [12] is an adaptation of the generalized Laplacian pyramid (GLP) method [49], which is an effective multiresolution analysis-based hyperspectral pansharpening method, for the fusion of HSIs and MSIs. Among the matrix factorization-based methods, we compared the CNMF method [24], vector total variation regularization method, HySure [28], and LTMR method [30]. Note that CNMF and HySure can also be used to address the fusion of HSIs and panchromatic (PAN) images directively. For tensor factorization-based methods, we considered the Super-resolution TEnsor-REcOnstruction (STEREO) method [33], Super-resolution based on COupled Tucker Tensor approximation (SCOTT) method [73], and Coupled Block term decompositions for hyperspectral Super-resoluTion with vARiability (CB-STAR) method [39], which also accounts for both spatially and spectrally localized changes while fusing. A MATLAB demo of the proposed GSFus method is available on the page https://github.com/FxyPd.

In GLPHS [12], *hypersharpening* [56] was used to adapt the GLP method [49] to effectively fuse HSIs and MSIs. Specifically, we assigned to each hyperspectral band a high resolution synthesized image obtained by a linear regression of MSI bands via least squares methods [12], [15], [56]. For CNMF and Hysure, we set the dimension of endmember to 30 as in [12], and we fine-tuned the λ_ϕ in Hysure to achieve the best fusion results. The subspace of LTMR was fixed to eight for fair comparisons with the proposed method, and we carefully tuned the regularization parameter λ. For STEREO, SCOTT, and CB-STAR, the parameters were set to be the same as in [39] for the Lake Tahoe data and the Ivanpah Playa data, and we hand tuned these parameters carefully on the other datasets. The other main parameters of the compared methods were set according to their original articles and codes for better fusion performances. For example, the λ_m in Hysure were set to 1 [28]. The number of clusters K and the path size $\sqrt{q}$ in LTMR were set to 200 and 7 [30], respectively.

2) Evaluation Indexes: In the experiments, we compared the experimental results of the proposed method with seven state-of-the-art fusion methods qualitatively and quantitatively. We used the following four indexes to thoroughly evaluate the quality of fusion results.

The first index chosen is the peak SNR (PSNR) [74]. PSNR of the ith band of an HSI is defined as

$$\text{PSNR}\left(\mathbf{Z}_i, \widehat{\mathbf{Z}}_i\right) = 10 \log_{10}\left(\frac{\max(\mathbf{Z}_i)^2}{\left\|\mathbf{Z}_i - \widehat{\mathbf{Z}}_i\right\|_F^2 / (WH)}\right) \quad (21)$$

where $\mathbf{Z}_i$ and $\widehat{\mathbf{Z}}_i$ denote the ith band of the true HR-HSI $\mathcal{Z}$ and its reconstruction $\widehat{\mathcal{Z}}$, respectively. Higher PSNR means better quality of the reconstruction.

The second index chosen is the spectral angle mapper (SAM) [17]. SAM measures the averaged angle between the pixel $\mathbf{z}_j$ in the ground truth and the reconstructed pixel $\widehat{\mathbf{z}}_j$ throughout the whole images. SAM is given in degrees and it can be expressed as

$$\text{SAM}\left(\mathcal{Z}, \widehat{\mathcal{Z}}\right) = \frac{1}{WH} \sum_{j=1}^{WH} \text{arcos} \frac{\widehat{\mathbf{z}}_j^T \mathbf{z}_j}{\|\widehat{\mathbf{z}}_j\|_2 \|\mathbf{z}_j\|_2}. \quad (22)$$

A smaller SAM is preferred as smaller SAM means less spectral distortions.

The third index is the relative dimensionless global error in synthesis (ERGAS) [29], which is defined as

$$\text{ERGAS}\left(\mathcal{Z}, \widehat{\mathcal{Z}}\right) = \frac{100}{d} \sqrt{\frac{1}{L} \sum_{i=1}^{L} \frac{\text{MSE}\left(\mathbf{Z}_i, \widehat{\mathbf{Z}}_i\right)}{\mu_{\widehat{\mathbf{Z}}_i}^2}} \quad (23)$$

where d is the spatial downsampling factor, $\text{MSE}(\mathbf{Z}_i, \widehat{\mathbf{Z}}_i)$ is the mean square error between $\mathbf{Z}_i$ and $\widehat{\mathbf{Z}}_i$, $\mu_{\widehat{\mathbf{Z}}_i}$ is the mean value of $\widehat{\mathbf{Z}}_i$. Smaller ERGAS means better reconstruction results.

The fourth index exploited is the universal image quality index (UIQI) [75]. In this article, we compute the UIQI for each band of HSI by first calculating UIQI within each sliding 32×32 window and averaging over all the windows in each bands. Then, we obtained the UIQI for the whole HSI by

averaging the UIQI of each bands. The UIQI for the band $\mathbf{Z}_i$ and $\widehat{\mathbf{Z}}_i$ is denoted as

$$\text{UIQI}\left(\mathbf{Z}_i, \widehat{\mathbf{Z}}_i\right) = \frac{1}{M} \sum_{j=1}^{M} \frac{\sigma_{\mathbf{Z}_j^i \widehat{\mathbf{Z}}_j^i}^2}{\sigma_{\mathbf{Z}_j^i} \sigma_{\widehat{\mathbf{Z}}_j^i}} \frac{2\mu_{\mathbf{Z}_j^i} \mu_{\widehat{\mathbf{Z}}_j^i}}{\mu_{\mathbf{Z}_j^i}^2 + \mu_{\widehat{\mathbf{Z}}_j^i}^2} \frac{2\sigma_{\mathbf{Z}_j^i} \sigma_{\widehat{\mathbf{Z}}_j^i}}{\sigma_{\mathbf{Z}_j^i}^2 + \sigma_{\widehat{\mathbf{Z}}_j^i}^2} \quad (24)$$

where M denotes the total number of windows, $\mathbf{Z}_j^i$ and $\widehat{\mathbf{Z}}_j^i$ are the jth window in the ith band of the true HSI and the reconstructed HSI, respectively. $\mu_{\mathbf{Z}_j^i}$, $\sigma_{\mathbf{Z}_j^i}$ are the mean and standard deviation of $\mathbf{Z}_j^i$, respectively. $\sigma_{\mathbf{Z}_j^i \widehat{\mathbf{Z}}_j^i}$ denotes the covariance of $\mathbf{Z}_j^i$ and $\widehat{\mathbf{Z}}_j^i$. The quality of the fusion results is better if its UIQI value is higher.

C. Experiments on Datasets With Localized Interimage Changes

In order to assess the fusion performance of the proposed method on datasets with localized changes between hyperspectral and MSIs, in this section, we perform experiments on three datasets that exhibit significant localized changes between hyperspectral and MSIs, namely the Indian Pines data with localized interimage changes, the Lake Tahoe data and the Ivanpah Playa data.

We presented the mean PSNR (MPSNR), SAM, ERGAS, and mean UIQI (MUIQI) of the recovered HSIs for the three datasets in Table I, where the best results were highlighted in bold. The running times are also presented in Table I. From Table I, it can be clearly seen that for the three datasets, the proposed method achieves considerable better overall reconstruction performance than the competitors in terms of the four evaluation indexes. The reason why the proposed method achieved much better fusion performance than the compared methods is that the proposed method introduced the $\ell_{2,1}$ norm to incorporate the interimage changes between hyperspectral and MSIs. Although CB-STAR also takes spatially and spectrally localized changes into consideration, our proposed method outperforms it on all the three datasets.

We presented the PSNR values of each restoration band for the three datasets in Fig. 7. It can be clearly observed from Fig. 7 that the PSNR values of the proposed method are significantly higher than that of compared methods at almost all the bands, which suggests that the proposed method achieves the best fusion results for all the datasets when there are localized changes between hyperspectral and MSIs.

For a visual comparison, the reconstructed images and corresponding error images of the tenth band for Indian Pines data with interimage changes are shown in Fig. 8. The error images were obtained by calculating the absolute values of the differencing images between ground truths and the estimated images. From Fig. 8, we can see that the reconstruction performance of the proposed method outperforms other competitors obviously. For the Indian Pines data, interimage changes occur in the area marked by the green box as shown in Fig. 6. We can see that the error image of the proposed method has fewer errors and is closer to the ground truth than other methods, in particular in the area of interimage changes. For the Lake Tahoe data, changes mainly occur at the crop circles

TABLE I

QUANTITATIVE INDEXES OF THE PROPOSED METHOD AND COMPARED METHODS ON DATASETS WITH LOCALIZED CHANGES

Data	Indexes	GLPHS	CNMF	HySure	STEREO	SCOTT	LTMR	CB-STAR	Proposed
Indian Pines with Inter-image changes	**MPSNR**	31.86	30.72	26.80	34.64	26.17	34.35	31.72	**36.84**
	SAM	2.61	1.46	3.58	1.36	3.56	1.35	2.29	**0.94**
	ERGAS	3.56	4.42	5.98	2.22	5.45	2.28	3.41	**1.73**
	MUIQI	0.95	0.95	0.92	0.98	0.93	0.97	0.96	**0.98**
	Time(s)	14.02	65.60	12.38	6.24	**1.76**	48.83	8.99	96.23
Lake Tahoe	**MPSNR**	26.99	21.58	19.17	27.18	25.23	25.25	30.60	**31.84**
	SAM	4.80	7.34	10.35	7.24	6.90	9.17	3.63	**2.35**
	ERGAS	5.52	10.24	11.48	5.65	6.49	8.39	3.53	**3.04**
	MUIQI	0.95	0.83	0.82	0.95	0.94	0.92	0.98	**0.99**
	Time(s)	4.41	4.99	3.23	3.64	**2.79**	25.64	14.33	34.11
Ivanpah Playa	**MPSNR**	28.95	26.87	23.35	28.52	26.74	29.22	32.93	**33.87**
	SAM	1.48	1.11	1.95	1.99	1.68	2.57	1.12	**0.82**
	ERGAS	2.52	3.14	4.53	2.66	3.16	2.98	1.57	**1.41**
	MUIQI	0.85	0.75	0.57	0.85	0.76	0.86	0.95	**0.97**
	Time(s)	5.21	29.84	3.22	0.72	**0.43**	28.81	17.52	47.88

TABLE II

QUANTITATIVE INDEXES OF THE PROPOSED METHOD AND COMPARED METHODS ON DATASETS WITHOUT LOCALIZED CHANGES

| Data	Indexes	GLPHS	CNMF	HySure	STEREO	SCOTT	LTMR	CB-STAR	Proposed
Indian Pines	**MPSNR**	33.26	34.16	27.47	36.15	26.48	36.05	32.33	**37.53**
	SAM	2.42	1.30	3.56	1.32	3.49	1.21	2.21	**0.92**
	ERGAS	2.87	2.33	5.59	1.98	5.19	1.84	3.13	**1.54**
	MUIQI	0.96	0.98	0.93	0.98	0.93	0.98	0.96	**0.98**
	Time(s)	33.02	30.27	44.62	6.04	**1.68**	48.15	8.83	96.74
Cuprite	**MPSNR**	41.95	42.98	40.95	42.71	36.89	45.11	42.95	**45.97**
	SAM	0.81	0.73	0.85	0.78	0.89	0.64	0.80	**0.55**
	ERGAS	1.01	0.92	1.16	0.97	1.52	0.79	1.05	**0.71**
	MUIQI	0.98	0.98	0.97	0.98	0.96	0.99	0.98	**0.99**
	Time(s)	8.19	35.47	12.37	3.57	**0.74**	33.26	3.35	95.90
Paris	**MPSNR**	29.90	28.66	27.27	29.81	29.85	30.48	30.31	**30.99**
	SAM	2.77	3.05	4.07	3.19	3.03	2.67	2.74	**2.51**
	ERGAS	5.52	6.42	7.58	5.53	5.57	5.14	5.25	**4.78**
	MUIQI	0.88	0.84	0.78	0.88	0.88	0.89	0.89	**0.91**
	Time(s)	2.44	12.56	11.21	3.26	**0.23**	17.54	1.42	25.98
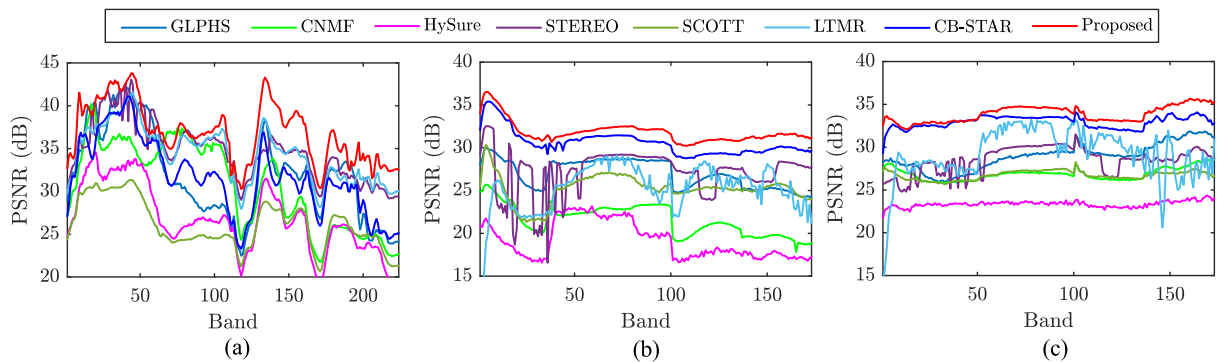
Fig. 7. Band-wise PSNR for three datasets. (a) Indian Pines with interimage changes. (b) Lake Tahoe. (c) Ivanpah Playa.

and around the lake, as shown in the last column of Fig. 4. The reconstructed images and corresponding error images of the tenth band of the Lake Tahoe data are shown in Fig. 9, where we can see that the reconstructed image of the proposed method looks closer to the ground truth. In terms of the error images, the error image of GLPHS is noisy. Many residuals remain in the error images of CNMF, HySure, STEREO,

SCOTT, and LTMR, in particular in the areas of changes, as shown in the last row of Fig. 4. Fewer residuals left in the error image of CB-STAR and the proposed GSFus method, highlighting the importance of taking interimages changes into consideration while fusing. For the Ivanpah Playa data, changes mainly occur in the central part and lower left corner of the images, as indicated in the last column of Fig. 4.
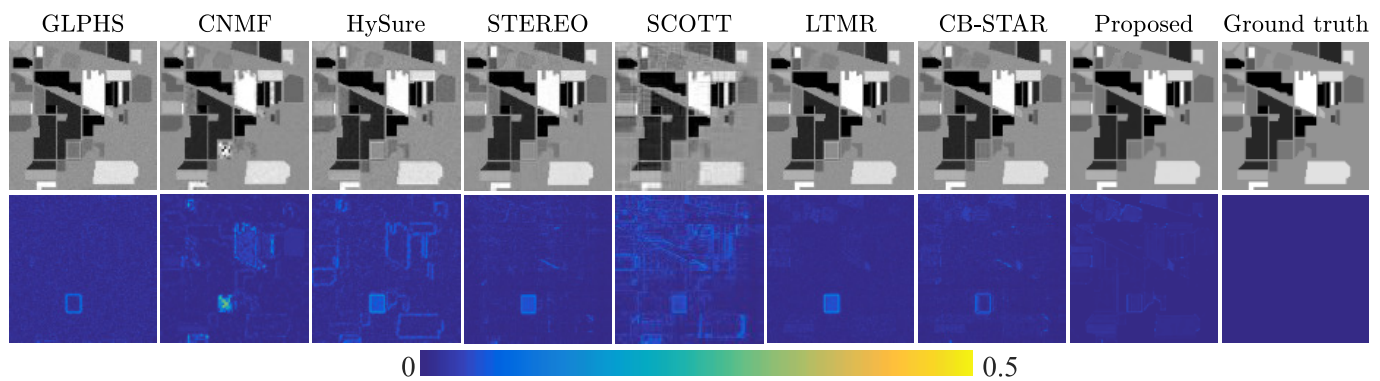
Fig. 8. Tenth band of reconstructed images (first row) and the corresponding error images (second row) for Indian Pines with changes data.
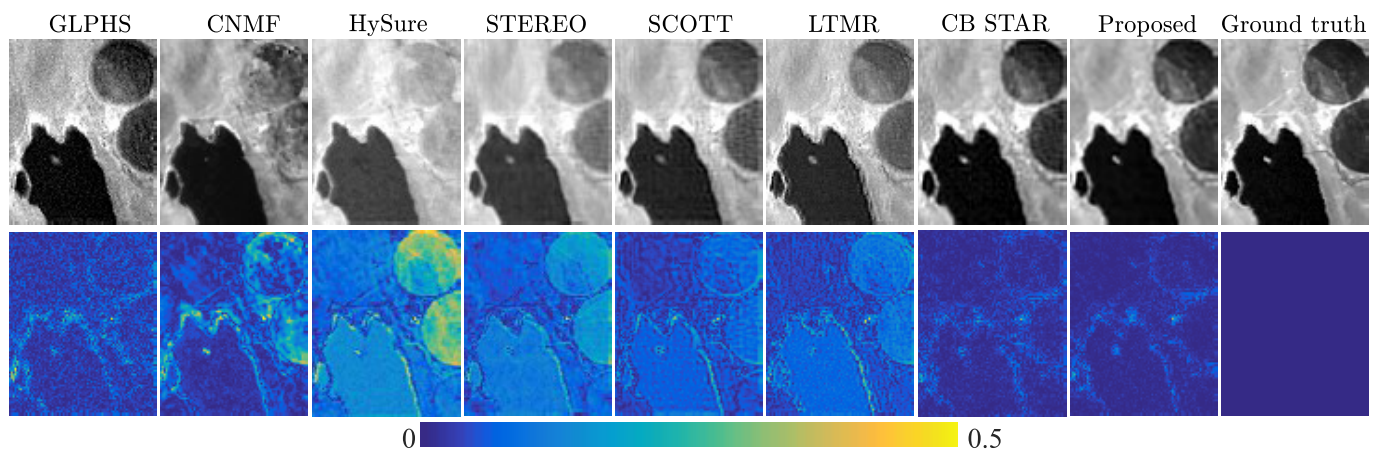

Fig. 9. Tenth band of reconstructed images (first row) and the corresponding error images (second row) for Lake Tahoe data.
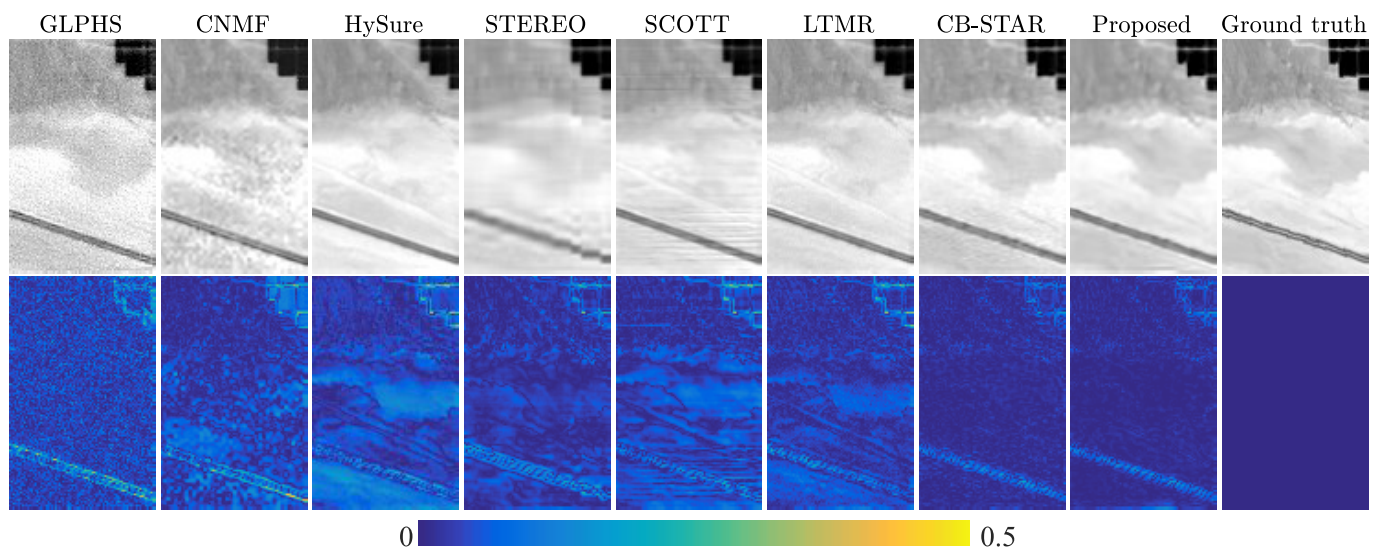

Fig. 10. 35th band of reconstructed images (first row) and the corresponding error images (second row) for Ivanpah Playa data.

We present the reconstructed image of the 35th band and its corresponding error images for the Ivanpah Playa data in Fig. 10. It is clear that the fusion result of the GLPHS is dissatisfactory as there is much noise left in the reconstructed image and error image. At the area where changes occur, there are many residuals left in the error images of CNMF, HySure, STEREO, SCOTT, and LTMR. The fusion results of CB-STAR and proposed GSFus method are much better than other methods. From the error images of CB-STAR and the proposed method on three datasets, it can be seen that fewer residuals left in the error images of the proposed method, suggesting that the proposed method is more effective than
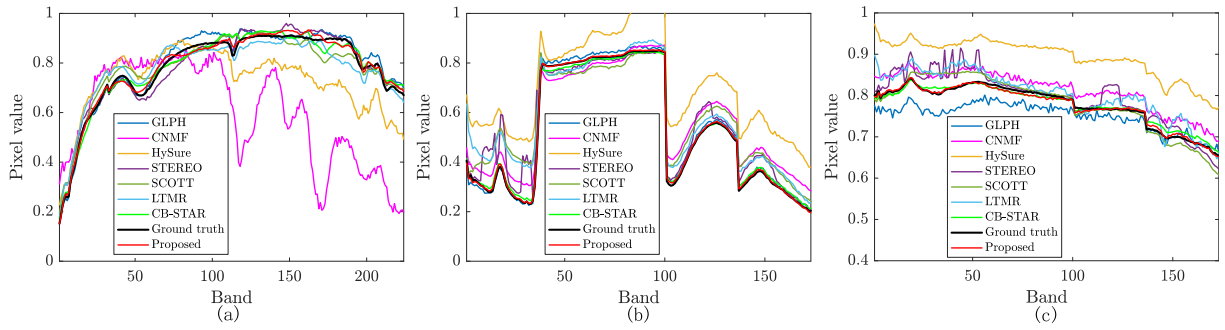
Fig. 11. Spectral vector of a selected pixel and the corresponding reconstruction results in (a) Indian Pines data with interimage changes, (b) Lake Tahoe data, and (c) Ivanpah Playa data.
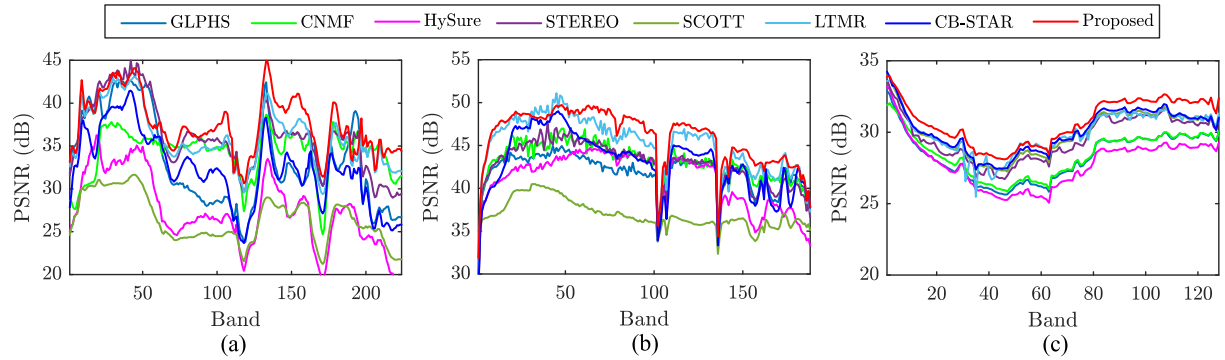

Fig. 12. Band-wise PSNR for three datasets. (a) Indian Pines. (b) Cuprite. (c) Paris.

CB-STAR, which is also a fusion method accounting for spatially and spectrally localized changes while fusing.

As the spectral signal is the most important information carried by HSIs to identify materials, we compare the spectral vectors of the reconstructed results in Fig. 11. As localized interimage changes occur between the observed LR-HSI and HR-MSI in these tested datasets, we can observe in Fig. 11 that the spectral vectors recovered by the fusion methods deviate significantly from the ground truth, except for CB-STAR and GSFus. Thanks to the introduction of $\ell_{2,1}$ norm to incorporate the interimage changes in the proposed GSFus method, the spectral vectors recovered by the proposed GSFus method are more closer to the ground truth and can retain critical spectral features than CB-STAR, which will be of great benefit to the further applications of the HSIs.

Considering the observations above, it is clear that the proposed method is effective in fusing hyperspectral and MSIs with localized changes between them and performs better than the other competitors on the three datasets overall.

D. Experiments on Datasets Without Obvious Localized Interimage Changes

In this section, we perform experiments on the datasets with no obvious localized changes between the hyperspectral and MSIs, namely the Indian Pines data, the Cuprite data, and the Paris data, to evaluate the fusion performance of the proposed GSFus method in comparison with other methods.

We summarized the MPSNR, SAM, ERGAS, and MUIQI of the recovered HSIs for the Indian Pines, Cuprite, and

Paris data in Table II. The running times are also presented in Table II. For better visualization, we marked the best results in bold. From Table II, we can see clearly that for the Indian Pines data, the fusion results of the proposed method are the best. STEREO, and LTMR get MPSNR values very close to the proposed method. For the Cuprite data, the performance LTMR is a little worse than the proposed method. For the Paris data, all the method achieves comparable results, and the proposed method is the best. In all, even on the datasets without obvious localized changes, the proposed method achieves the best overall reconstruction performance in terms of the four evaluation indexes, which suggests that the proposed method is a more flexible and robust fusion method. GLPHS method is a multiresolution analysis-based pansharpening method adapted for fusing HSI with MSI and is prone to be affected by noise. CNMF is spectral unmixing-based fusion method, thus it can obtain relatively good results under the impact of Gaussian noise. However, it does not impose any priors on the abundance, which, to some degree, limited their fusion performance. STEREO, SCOTT, CB-STAR are state-of-the-art tensor factorization-based methods. STEREO and CB-STAR can obtain relatively good fusion results on all the datasets, but SCOTT failed to achieve reasonable results on Indian Pines and Cuprite data. HySure, LTMR, and the proposed fusion method are subspace-based method. One of the differences between the three methods is the priors imposed to regularize the subspace coefficients. HySure and LTMR use vector total variation and LTMR as the prior of the subspace coefficients, respectively, whereas the proposed
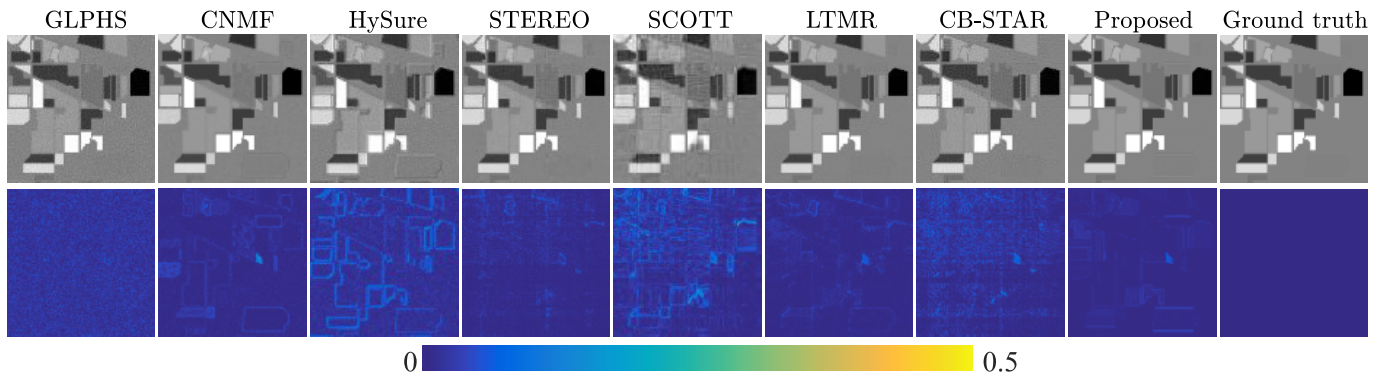
Fig. 13.   155th band of reconstructed images (first row) and the corresponding error images (second row) for Indian Pines data.
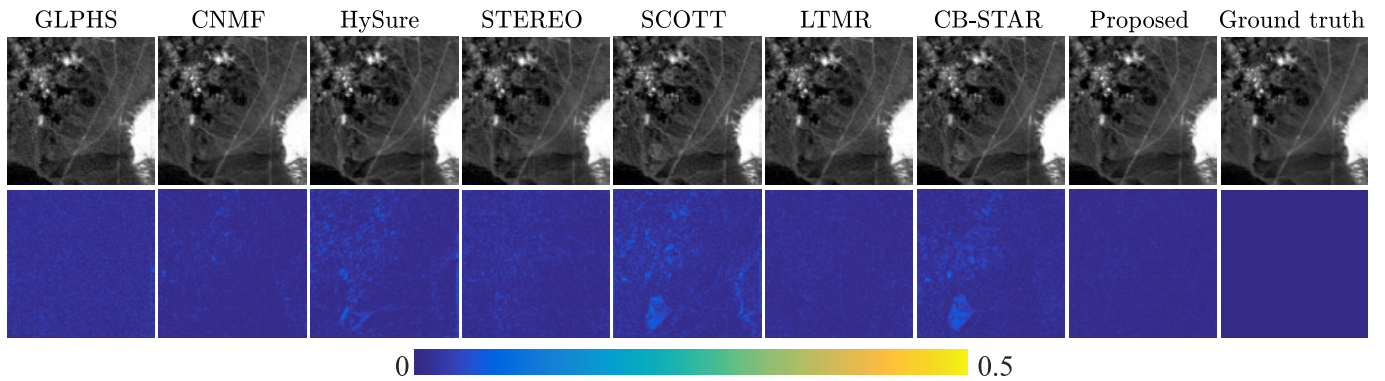

Fig. 14.   170th band of reconstructed images (first row) and the corresponding error images (second row) for Cuprite data.
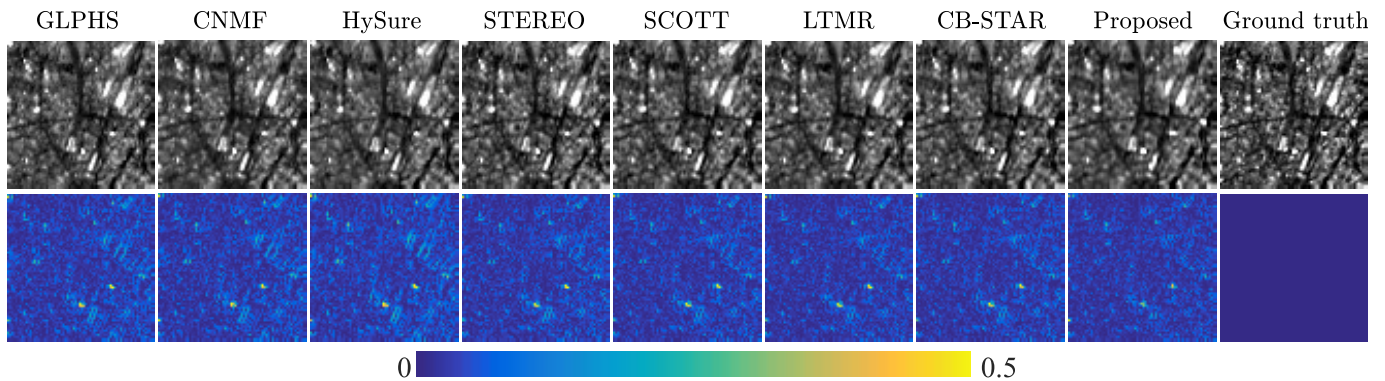

Fig. 15.   120th band of reconstructed images (first row) and the corresponding error images (second row) for Paris data.

method implicitly regularizes the subspace coefficients using a state-of-the-art denoiser, BM3D, which can achieve satisfying denoising performance while preserving the image details. The reason why the proposed method achieved the best overall fusion results partly boils down to the implicit prior, i.e., BM3D, which denoises the subspace coefficients by fully making use of the sparsity and the nonlocal self-similarity of the subspace coefficients. In terms of the running time, from Tables I and II, it can be seen that SCOTT takes the shortest time to complete fusion on all the datasets. The proposed method achieves the best performance at the cost of increased computation time.

The PSNR values of each restoration band for simulated Indian Pines data, Cuprite data, and the Paris data are depicted in Fig. 12. It can be clearly seen from Fig. 12 that the PSNRs of the proposed method are higher than that of compared methods at most of the bands, which suggests that the proposed method achieved the best reconstructed fusion results for the three datasets overall.

For a visual comparison, Figs. 13–15 display the reconstructed images and corresponding error images of the 155th band for Indian Pines data, the 170th band for Cuprite data, and the 120th band for Paris data, respectively. It can be seen that the reconstructed images of the proposed method are more closer to the ground truths, and the residuals left in the error images are fewer than the other methods. The reason that the proposed method performs better than other subspace-based methods, i.e., HySure and LTMR, can be attributed to the

effective denoiser used as the prior of subspace coefficients in the proposed method.

Taking all the above observations into consideration, it is clear that the proposed method is not only effective in fusing hyperspectral and MSIs with localized changes between them, but also performs better than the other compared methods when codes with the images without obvious localized change. The reason is that the proposed method introduces the $\ell_{2,1}$ norm to incorporate the interimage changes between hyperspectral and MSIs and uses an effective denoiser as the prior of subspace coefficients, which make the proposed method a more flexible and robust fusion method.

E. Ablation Study

To verify the effectiveness of the contributions involved in the proposed GSFus model, we present a series of ablation study by comparing the MPSNR of the recovered HSIs in this section. First, we verify the effectiveness of the plugged in prior as the regularizer on the subspace coefficients $\mathbf{A}$ by setting the β to 0, such that the objective function will be reduced to the following equation:

$$\min_{\mathbf{A}} \frac{1}{2}\|\mathbf{Y}_h - \mathbf{DABS}\|_F^2 + \frac{\lambda}{2}\|\mathbf{Y}_m - \mathbf{RDA}\|_{2,1}. \quad (25)$$

Second, to verify the effectiveness of the proposed $\ell_{2,1}$ norm on the residual term of the MSI to account for the localized changes occurring between the hyperspectral and MSIs, we replaced the $\ell_{2,1}$ norm on the reconstruction error of the MSI of the objective function with Frobenius norm to obtain as follows:

$$\min_{\mathbf{A}} \frac{1}{2}\|\mathbf{Y}_h - \mathbf{DABS}\|_F^2 + \frac{\lambda}{2}\|\mathbf{Y}_m - \mathbf{RDA}\|_F^2 + \beta\phi(\mathbf{A}). \quad (26)$$

Then, we fused the HSIs and MSIs using (25) and (26), respectively. The corresponding reconstruction MPSNRs are given in Table III. Besides, we also listed the recovered MPSNRs of the proposed GSFus model as a reference. The best results are marked in bold for better visualization. From the first row and the third row of Table III, we can clearly see that the MPSNRs of (25) is significantly smaller than the proposed GSFus method, which means that the plugged in denoiser in the GSFus method is able to effectively regularize the subspace coefficients, $\mathbf{A}$, and can help to improve the fusion performance. By comparing the second row and the third row of Table III, we can observe that the proposed GSFus method could always get the best MPSNRs on all the datasets. Equation (26) could get comparable reconstructed MPSNRs on images without obvious interimage changes. But for images with significant interimage changes, the reconstructed MPSNRs of (26) are significantly smaller than the proposed GSFus. This is because the use of $\ell_{2,1}$ norm in the proposed GSFus method, instead of Frobenius norm, promotes group sparsity and thus enables us to characterize the spatial and spectral changes between observed MSI, $\mathbf{Y}_m$, and the underlying HR-HSI, $\mathcal{Z}$.

To further demonstrate the advantage of the proposed $\ell_{2,1}$ norm regularization on the residual term of the MSI, we utilized a state-of-the-art low-rank tensor decomposition, i.e., canonical polyadic decomposition (CPD), to model the HSI and MSI degradation and then fused the degraded images using the STEREO method and the corresponding modified methods. According to [33], the HR-HSI, $\mathcal{Z}$, can be represented as $\mathcal{Z} = [\![\mathbf{I}, \mathbf{J}, \mathbf{K}]\!]$, where $\mathbf{I} \in \mathbb{R}^{W \times F}$, $\mathbf{J} \in \mathbb{R}^{H \times F}$ and $\mathbf{K} \in \mathbb{R}^{L \times F}$ are called the low-rank latent factors of the third-order tensor, respectively, and F is the tensor rank. The objective function of STEREO is given as

$$\min_{\mathbf{I},\mathbf{J},\mathbf{K}} \|\mathcal{Y}_h - [\![\mathbf{P}_1\mathbf{I}, \mathbf{P}_2\mathbf{J}, \mathbf{K}]\!]\|_F^2 + \gamma \|\mathcal{Y}_m - [\![\mathbf{I}, \mathbf{J}, \mathbf{P}_M\mathbf{K}]\!]\|_F^2 \quad (27)$$

where $\mathbf{P}_1 \in \mathbb{R}^{w \times W}$, $\mathbf{P}_2 \in \mathbb{R}^{h \times H}$, and $\mathbf{P}_M \in \mathbb{R}^{l \times L}$ are the downsampling matrices along the width, height, and spectral modes, respectively. $\gamma > 0$ is a preselected parameter that weights the importance of each image in estimating $\mathbf{I}$, $\mathbf{J}$, and $\mathbf{K}$. To verify the effectiveness of the proposed $\ell_{2,1}$ norm to regularize the residual of the MSI image, we replace the Frobenius norm with $\ell_{2,1}$ norm on the residual term of the MSI in (27) to form a new objective function as

$$\min_{\mathbf{I},\mathbf{J},\mathbf{K}} \|\mathcal{Y}_h - [\![\mathbf{P}_1\mathbf{I}, \mathbf{P}_2\mathbf{J}, \mathbf{K}]\!]\|_F^2 + \gamma \|\mathcal{Y}_m - [\![\mathbf{I}, \mathbf{J}, \mathbf{P}_M\mathbf{K}]\!]\|_{2,1}. \quad (28)$$

Then, we fused the HSIs and MSIs using the STEREO method, i.e., (27), and (28), respectively. The reconstructed MPSNRs are presented in Table IV and the best results are marked in bold for better visualization. It can be seen from Table IV that (28) could achieve good fusion results on all the images. STEREO obtains reconstructed MPSNRs very close to (28) on images without obvious interimage changes. But for images with significant interimage changes, the reconstructed MPSNRs of STEREO are significantly smaller than (28). The experimental results prove the superiority of using $\ell_{2,1}$ norm to regularize the residual term of the MSI.

F. Parameters Analysis

The proposed method includes three parameters, namely, the dimension of subspace, k, and the tradeoff parameters, λ and β. To assess the sensitivity of the proposed GSFus method to the parameters, we performed experiments on all the datasets with various settings of k, λ, and β. The reconstructed MPSNRs of the proposed method as a function of k, λ, and β are presented in Fig. 16. Fig. 16(a) presents the fusion performance of the proposed method as a function of k. It can be seen from the figure that the MPSNR values of all the datasets are practically constant when the dimension of subspace is greater than 8. It implies that the proposed method is extremely robust to the overestimation of subspace dimension. In this article, we fixed the dimension of subspace to eight on all the datasets.

The parameters, λ and β, determine the intensity of the group sparsity regularization on $(\mathbf{Y}_m - \mathbf{RDA})$ and the implicit regularization on $\mathbf{A}$, respectively. These two parameters are important to the fusion of hyperspectral and MSIs. Fig. 16(b) presents the fusion performance of the proposed method with the change of λ. Ideally, the determination of parameter λ
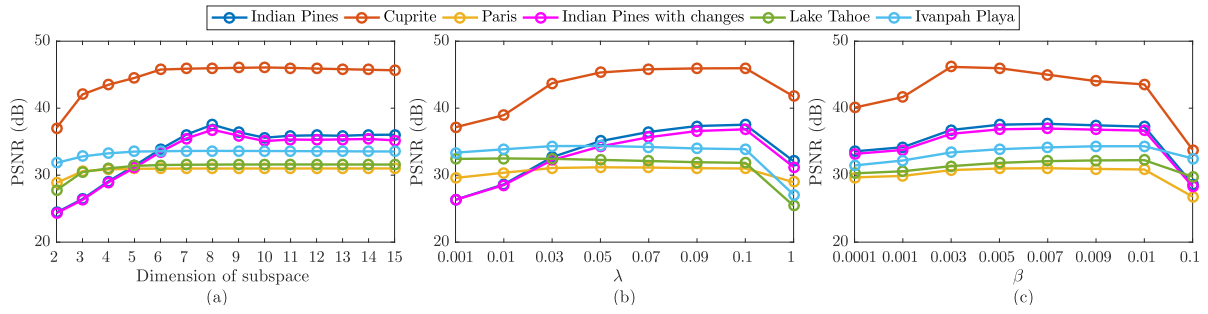
Fig. 16. MPSNR of the proposed method with the changes of dimension of subspace k, λ, and β. (a) k. (b) λ. (c) β.
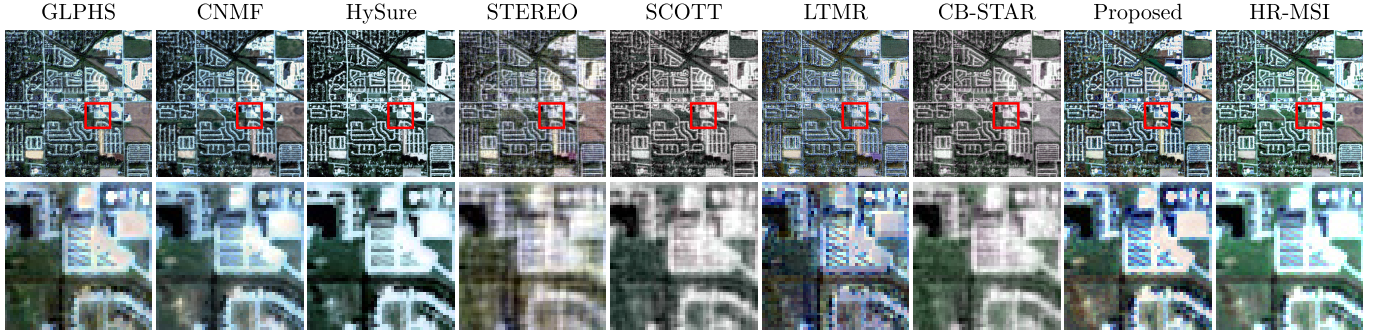

Fig. 17. Fusion results for full resolution dataset. The first row presents the fusion images composed of bands 20, 10, and 1. The second row shows the enlarged area marked by red rectangle in the first row.

TABLE III

RECONSTRUCTED MPSNR VALUES OF (25), (26), AND THE PROPOSED GSFus METHOD

Methods	Images without obvious inter-image changes			Images with inter-image changes		
	Indian Pines	Cuprite	Paris	Indian Pines with inter-image changes	Lake Tahoe	Ivanpah Playa
Eq.(25)	33.54	40.07	29.66	33.17	30.33	31.83
Eq.(26)	37.41	45.83	30.55	35.13	27.95	32.01
GSFus	**37.53**	**45.97**	**30.99**	**36.84**	**31.84**	**33.87**

TABLE IV

RECONSTRUCTED MPSNR VALUES OF STEREO METHOD AND (28)

Methods	Images without obvious inter-image changes			Images with inter-image changes		
	Indian Pines	Cuprite	Paris	Indian Pines with inter-image changes	Lake Tahoe	Ivanpah Playa
STEREO	36.15	**42.71**	29.81	34.61	27.18	28.52
Eq.(28)	**36.29**	42.64	**30.49**	**35.78**	**29.36**	**29.73**

should depend on the intensity of changes occurring between the acquisition of HSI and MSI. However, it is not practical to determine the intensity of changes in real situations. From Fig. 16(b), we can see that the MPSNRs of the Indian Pines and Cuprite data increase gradually when λ varies from 0.01 to 0.1, whereas the MPSNRs of the other dataset remain relatively stable. By setting λ to 0.1, we could obtain good results for all the datasets. Parameter β controls the intensity of denoising the subspace coefficients. From Fig. 16(c), we can see that except for Cuprite data, MPSNRs of the other five datasets remain relatively stable when β varies from

0.003 to 0.01, whereas the MPSNR values of Cuprite data show a little decrease. By setting β to 0.005, we can obtain desired fusion results for all the datasets.

G. Experiment on Full Resolution Dataset

To further evaluate the effectiveness of the proposed method, an experiment on full resolution data was carried out using an HSI captured by the Hyperion sensor and an MSI obtained by the Sentinel-2A satellite. The spectral response function was estimated using the method in [28] and we simply

adopted a 3×3 average kernel for the PSF. We show the fusion results in Fig. 17. As shown in the figure, all the methods can improve the spatial resolution of the Hyperion LR-HSI to a different extent. The results of STEREO, SCOTT, and CB-STAR look blurry compared with the others. There are some artificial effects in the result of LTMR. GLPHS, CNMF, HySure, and the proposed GSFus method can recover more spatial details in the fusion results. The experiment on full-resolution dataset demonstrates that the proposed GSFus method can perform well in the fusion of full-resolution dataset.

V. CONCLUSION

In this article, we proposed a novel GSFus method to fuse hyperspectral and MSIs based on matrix factorization. By imposing the $\ell_{2,1}$ norm on the reconstruction error of the MSI term to promote the group sparsity, the proposed fusion method can account for the case in which localized changes occur between the hyperspectral and multispectra images obtained from different time or sensors. Meanwhile, we propose to use a plugged in prior for the subspace coefficients by exploiting the plug-and-play framework, which allows us to insert a state-of-the-art denoiser, namely BM3D, to regularize the subspace coefficients. Two kinds of dataset, i.e., datasets with/without interimage changes, and a full resolution dataset were used to evaluate the effectiveness of the proposed method. We performed experiments on the datasets by comparing the proposed fusion methods with seven state-of-the-art competitors on the experimental datasets according to four indexes, i.e., PSNR, SAM, ERGAS, and UIQI, to evaluate the effectiveness of the proposed method. Experimental results suggest that the proposed method is effective on all the datasets and can obtain better fusion results than the competitors.

REFERENCES

[1] J. M. Bioucas-Dias et al., "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 5, no. 2, pp. 354–379, Apr. 2012.

[2] L. Zhuang and M. K. Ng, "Hyperspectral mixed noise removal by ℓ_1-norm-based subspace representation," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 13, pp. 1143–1157, 2020.

[3] X. Lu, Y. Yuan, and X. Zheng, "Joint dictionary learning for multispectral change detection," IEEE Trans. Cybern., vol. 47, no. 4, pp. 884–897, Apr. 2017.

[4] L. Zhuang, C. Lin, M. A. T. Figueiredo, and J. M. Bioucas-Dias, "Regularization parameter selection in minimum volume hyperspectral, unmixing," IEEE Trans. Geosci. Remote Sens., vol. 57, no. 12, pp. 9858–9877, Dec. 2019.

[5] L. Gao, Z. Wang, L. Zhuang, H. Yu, B. Zhang, and J. Chanussot, "Using low-rank representation of abundance maps and nonnegative tensor factorization for hyperspectral nonlinear unmixing," IEEE Trans. Geosci. Remote Sens., early access, Mar. 25, 2021, doi: 0.1109/TGRS.2021.3065990.

[6] M. T. Eismann, J. Meola, and R. C. Hardie, "Hyperspectral change detection in the presenceof diurnal and seasonal variations," IEEE Trans. Geosci. Remote Sens., vol. 46, no. 1, pp. 237–249, Jan. 2008.

[7] L. Zhuang, L. Gao, B. Zhang, X. Fu, and J. M. Bioucas-Dias, "Hyperspectral image denoising and anomaly detection based on low-rank and sparse representations," IEEE Trans. Geosci. Remote Sens., early access, Dec. 9, 2020, doi: 10.1109/TGRS.2020.3040221.

[8] X. Fu, S. Jia, L. Zhuang, M. Xu, J. Zhou, and Q. Li, "Hyperspectral anomaly detection via deep plug-and-play denoising CNN regularization," IEEE Trans. Geosci. Remote Sens., vol. 59, no. 11, pp. 9553–9568, Jan. 2021.

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

[10] L. Zhuang, X. Fu, M. K. Ng, and J. M. Bioucas-Dias, "Hyperspectral image denoising based on global and nonlocal low-rank factorizations," IEEE Trans. Geosci. Remote Sens., early access, Jan. 8, 2021, doi: 10.1109/TGRS.2020.3046038.

[11] R. Dian, S. Li, L. Fang, and Q. Wei, "Multispectral and hyperspectral image fusion with spatial–spectral sparse representation," Inf. Fusion, vol. 49, pp. 262–270, Sep. 2019.

[12] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," IEEE Geosci. Remote Sens. Mag., vol. 5, no. 2, pp. 29–56, Jun. 2017.

[13] F. Gao, J. Masek, M. Schwaller, and F. Hall, "On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance," IEEE Trans. Geosci. Remote Sens., vol. 44, no. 8, pp. 2207–2218, Aug. 2006.

[14] I. V. Emelyanova, T. R. McVicar, T. G. Van Niel, L. T. Li, and A. I. J. M. van Dijk, "Assessing the accuracy of blending Landsat–MODIS surface reflectances in two landscapes with contrasting spatial and temporal dynamics: A framework for algorithm selection," Remote Sens. Environ., vol. 133, pp. 193–209, Jun. 2013.

[15] R. Restaino, G. Vivone, P. Addesso, and J. Chanussot, "Hyperspectral sharpening approaches using satellite multiplatform data," IEEE Trans. Geosci. Remote Sens., vol. 59, no. 1, pp. 578–596, Jan. 2021.

[16] L. Loncan et al., "Hyperspectral pansharpening: A review," IEEE Trans. Geosci. Remote Sens., vol. 3, no. 3, pp. 27–46, Sep. 2015.

[17] Y. Zhang, S. De Backer, and P. Scheunders, "Noise-resistant wavelet-based Bayesian fusion of multispectral and hyperspectral images," IEEE Trans. Geosci. Remote Sens., vol. 47, no. 11, pp. 3834–3843, Nov. 2009.

[18] R. A. Borsoi, T. Imbiriba, and J. C. M. Bermudez, "Super-resolution for hyperspectral and multispectral image fusion accounting for seasonal spectral variability," IEEE Trans. Image Process., vol. 29, pp. 116–127, 2020.

[19] R. Borsoi et al., "Spectral variability in hyperspectral data unmixing: A comprehensive review," IEEE Geosci. Remote Sens. Mag., early access, May 21, 2021, doi: 10.1109/MGRS.2021.3071158.

[20] A. Zare and K. Ho, "Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing," IEEE Signal Process. Mag., vol. 31, no. 1, pp. 95–104, Jan. 2014.

[21] B. Somers, G. P. Asner, L. Tits, and P. Coppin, "Endmember variability in spectral mixture analysis: A review," Remote Sens. Environ., vol. 115, no. 7, pp. 1603–1616, 2011.

[22] R. Kawakami, Y. Matsushita, J. Wright, M. Ben-Ezra, Y.-W. Tai, and K. Ikeuchi, "High-resolution hyperspectral imaging via matrix factorization," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2011, pp. 2329–2336.

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

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

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

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

[27] R. C. Hardie, M. T. Eismann, and G. L. Wilson, "MAP estimation for hyperspectral image resolution enhancement using an auxiliary sensor," IEEE Trans. Image Process., vol. 13, no. 9, pp. 1174–1184, Sep. 2004.

[28] M. Simoes, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," IEEE Trans. Geosci. Remote Sens., vol. 53, no. 6, pp. 3373–3388, Jun. 2015.

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

[30] R. Dian and S. Li, "Hyperspectral image super-resolution via subspace-based low tensor multi-rank regularization," *IEEE Trans. Image Process.*, vol. 28, no. 10, pp. 5135–5146, Oct. 2019.

[31] K. Zhang, M. Wang, S. Yang, and L. Jiao, "Spatial–spectral-graph-regularized low-rank tensor decomposition for multispectral and hyperspectral image fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1030–1040, Apr. 2018.

[32] S. Li, R. Dian, L. Fang, and J. M. Bioucas-Dias, "Fusing hyperspectral and multispectral images via coupled sparse tensor factorization," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 4118–4130, Aug. 2018.

[33] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, and W.-K. Ma, "Hyperspectral super-resolution: A coupled tensor factorization approach," *IEEE Trans. Signal Process.*, vol. 66, no. 24, pp. 6503–6517, Dec. 2018.

[34] 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.

[35] 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.

[36] Y. Chang, L. Yan, X. Zhao, H. Fang, Z. Zhang, and S. Zhong, "Weighted low-rank tensor recovery for hyperspectral image restoration," *IEEE Trans. Cybern.*, vol. 50, no. 11, pp. 4558–4572, Nov. 2020.

[37] Y. Bu *et al.*, "Hyperspectral and multispectral image fusion via graph Laplacian-guided coupled tensor decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 648–662, Jan. 2021.

[38] W. He, Y. Chen, N. Yokoya, C. Li, and Q. Zhao, "Hyperspectral super-resolution via coupled tensor ring factorization," *Pattern Recognit.*, vol. 122, Feb. 2022, Art. no. 108280.

[39] R. A. Borsoi, C. Prevost, K. Usevich, D. Brie, J. C. M. Bermudez, and C. Richard, "Coupled tensor decomposition for hyperspectral and multispectral image fusion with inter-image variability," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 3, pp. 702–717, Jan. 2021.

[40] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Multispectral and hyperspectral image fusion using a 3-D-convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 639–643, May 2017.

[41] R. Dian, S. Li, A. Guo, and L. Fang, "Deep hyperspectral image sharpening," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5345–5355, Nov. 2018.

[42] F. Zhou, R. Hang, Q. Liu, and X. Yuan, "Pyramid fully convolutional network for hyperspectral and multispectral image fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 5, pp. 1549–1558, May 2019.

[43] X. Han, J. Yu, J. Luo, and W. Sun, "Hyperspectral and multispectral image fusion using cluster-based multi-branch BP neural networks," *Remote Sens.*, vol. 11, no. 10, p. 1173, May 2019.

[44] J. Yang, Y.-Q. Zhao, and J. Chan, "Hyperspectral and multispectral image fusion via deep two-branches convolutional neural network," *Remote Sens.*, vol. 10, no. 5, p. 800, May 2018.

[45] J. Xue, Y.-Q. Zhao, Y. Bu, W. Liao, J. C.-W. Chan, and W. Philips, "Spatial–spectral structured sparse low-rank representation for hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 30, pp. 3084–3097, 2021.

[46] W. J. Carper, T. M. Lillesand, and R. W. 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.

[47] B. Aiazzi, S. Baronti, and M. Selva, "Improving component substitution pansharpening through multivariate regression of MS+Pan data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3230–3239, Oct. 2007.

[48] J. G. Liu, "Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details," *Int. J. Remote Sens.*, vol. 21, no. 18, pp. 3461–3472, Nov. 2000.

[49] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "MTF-tailored multiscale fusion of high-resolution MS and PAN imagery," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 5, pp. 591–596, May 2006.

[50] 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.

[51] X. X. Zhu, C. Grohnfeld, 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.

[52] R. B. Gomez, A. Jazaeri, and M. Kafatos, "Wavelet-based hyperspectral and multispectral image fusion," *Proc. SPIE*, vol. 4383, pp. 36–42, Jun. 2001.

[53] B. Aiazzi, L. Alparone, L. Santurri, and M. Selva, "Spatial resolution enhancement of ASTER thermal bands," *Proc. SPIE*, vol. 5982, Oct. 2005, Art. no. 59821G.

[54] D. Picone, R. Restaino, G. Vivone, P. Addesso, M. D. Mura, and J. Chanussot, "Band assignment approaches for hyperspectral sharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 739–743, May 2017.

[55] Z. Chen, H. Pu, B. Wang, and G.-M. Jiang, "Fusion of hyperspectral and multispectral images: A novel framework based on generalization of pan-sharpening methods," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 8, pp. 1418–1422, Aug. 2014.

[56] M. Selva, B. Aiazzi, F. Butera, L. Chiarantini, and S. Baronti, "Hypersharpening: A first approach on SIM-GA data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 3008–3024, Jun. 2015.

[57] Y. Chen, T.-Z. Huang, L.-J. Deng, X.-L. Zhao, and M. Wang, "Group sparsity based regularization model for remote sensing image stripe noise removal," *Neurocomputing*, vol. 267, pp. 95–106, Dec. 2017.

[58] Y. Chen, W. He, N. Yokoya, and T.-Z. Huang, "Hyperspectral image restoration using weighted group sparsity-regularized low-rank tensor decomposition," *IEEE Trans. Cybern.*, vol. 50, no. 8, pp. 3556–3570, Aug. 2020.

[59] L. Zhuang, M. K. Ng, X. Fu, and J. M. Bioucas-Dias, "Hy-demosaicing: Hyperspectral blind reconstruction from spectral subsampling," *IEEE Trans. Geosci. Remote Sens.*, early access, Aug. 13, 2021, doi: 10.1109/TGRS.2021.3102136.

[60] L. Zhuang and J. M. Bioucas-Dias, "Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 730–742, Mar. 2018.

[61] L. Zhuang and M. K. Ng, "FastHyMix: Fast and parameter-free hyperspectral image mixed noise removal," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Sep. 29, 2021, doi: 10.1109/TNNLS.2021.3112577.

[62] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[63] Q. Wei, N. Dobigeon, and J. Tourneret, "Fast fusion of multi-band images based on solving a Sylvester equation," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4109–4121, Nov. 2015.

[64] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.

[65] T.-X. Jiang, L. Zhuang, T.-Z. Huang, X.-L. Zhao, and J. M. Bioucas-Dias, "Adaptive hyperspectral mixed noise removal," *IEEE Trans. Geosci. Remote Sens.*, early access, Jun. 15, 2021, doi: 10.1109/TGRS.2021.3085779.

[66] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2862–2869.

[67] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, Sep. 2018.

[68] W. He, H. Zhang, L. Zhang, and H. Shen, "Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 178–188, Jan. 2016.

[69] 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.

[70] G. Vivone *et al.*, "Pansharpening based on semiblind deconvolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 1997–2010, Apr. 2015.

[71] M. Selva, L. Santurri, and S. Baronti, "Improving hypersharpening for worldview-3 data," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 6, pp. 987–991, Jun. 2019.

[72] R. Dian, S. Li, and X. Kang, "Regularizing hyperspectral and multispectral image fusion by CNN denoiser," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 3, pp. 1124–1135, Mar. 2021.

[73] C. Prevost, K. Usevich, P. Comon, and D. Brie, "Hyperspectral super-resolution with coupled tucker approximation: Recoverability and SVD-based algorithms," *IEEE Trans. Signal Process.*, vol. 68, pp. 931–946, 2020.

[74] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[75] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.

Xiyou Fu (Member, IEEE) received the bachelor's degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2012, the M.S. degree in electronic and communication engineering and the Ph.D. degree in cartography and geography information system from the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China, in 2015 and 2019, respectively.

He is currently a Post-Doctoral Researcher with Shenzhen University, Shenzhen, China. His research interests include hyperspectral image restoration, anomaly detection, and super-resolution.
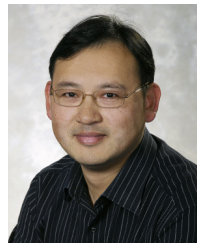
Sen Jia (Senior Member, IEEE) received the B.E. and Ph.D. degrees from the College of Computer Science, Zhejiang University, Hangzhou, China, in 2002 and 2007, respectively.

Since 2008, he has been with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China, where he is currently a Full Professor. His research interests include hyperspectral image processing, signal and image processing, and machine learning.

Meng Xu (Member, IEEE) received the B.S. and M.E. degrees in electrical engineering from the Ocean University of China, Qingdao, China, in 2011 and 2013, respectively, and the Ph.D. degree from the University of New South Wales, Canberra, ACT, Australia, in 2017.

She is an Assistant Professor with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China. Her research interests include cloud removal and remote sensing image processing.

Jun Zhou (Senior Member, IEEE) received the B.S. degree in computer science and the B.E. degree in international business from the Nanjing University of Science and Technology, Nanjing, China, in 1996 and 1998, respectively, the M.S. degree in computer science from Concordia University, Montreal, QC, Canada, in 2002, and the Ph.D. degree in computing science from the University of Alberta, Edmonton, AB, Canada, in 2006.

He was a Research Fellow with the Research School of Computer Science, The Australian National University, Canberra, ACT, Australia, and a Researcher with the Canberra Research Laboratory, National Information and Communications Technology Australia, Canberra, ACT. In 2012, he joined the School of Information and Communication Technology, Griffith University, Nathan, QLD, Australia, where he is currently an Associate Professor. His research interests include pattern recognition, computer vision, and spectral imaging and their applications in remote sensing and environmental informatics.

Qingquan Li received the M.S. degree in engineering and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1988 and 1998, respectively.

From 1988 to 1996, he was an Assistant Professor with Wuhan University, where he was an Associate Professor, from 1996 to 1998. He has been a Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, since 1998.

He is the President of Shenzhen University, Shenzhen, China, where he is the Director of the Shenzhen Key Laboratory of Spatial Information Smart Sensing and Services. His research interests include photogrammetry, remote sensing, and intelligent transportation systems.

Dr. Li is an expert in modern traffic with the National 863 Plan and an Editorial Board Member of the *Surveying and Mapping Journal* and the *Wuhan University Journal-Information Science Edition*.