Hyperspectral Image Superresolution Using Global Gradient Sparse and Nonlocal Low-Rank Tensor Decomposition With Hyper-Laplacian Prior

Yidong Peng, Weisheng Li, Xiaobo Luo, and Jiao Du

***Abstract*—This article presents a novel global gradient sparse and nonlocal low-rank tensor decomposition model with a hyper-Laplacian prior for hyperspectral image (HSI) superresolution to produce a high-resolution HSI (HR-HSI) by fusing a low-resolution HSI (LR-HSI) with an HR multispectral image (HR-MSI). Inspired by the investigated hyper-Laplacian distribution of the gradients of the difference images between the upsampled LR-HSI and latent HR-HSI, we formulate the relationship between these two datasets as a ℓ_p $(0 < p < 1)$-norm term to enforce spectral preservation. Then, the relationship between the HR-MSI and latent HR-HSI is built using a tensor-based fidelity term to recover the spatial details. To effectively capture the high spatio-spectral-nonlocal similarities of the latent HR-HSI, we design a novel nonlocal low-rank Tucker decomposition to model the 3-D regular tensors constructed from the grouped nonlocal similar HR-HSI cubes. The global spatial-spectral total variation regularization is then adopted to ensure the global spatial piecewise smoothness and spectral consistency of the reconstructed HR-HSI from nonlocal low-rank cubes. Finally, an alternating direction method of multipliers-based algorithm is designed to efficiently solve the optimization problem. Experiments on both the synthetic and real datasets collected by different sensors show the effectiveness of the proposed method, from visual and quantitative assessments.**

***Index Terms*—Global gradient sparse, hyper-Laplacian, hyperspectral image, nonlocal low-rank, superresolution, total variation, tucker decomposition.**

Manuscript received January 1, 2021; revised February 22, 2021; accepted April 24, 2021. Date of publication April 28, 2021; date of current version June 8, 2021. This work was supported in part by the Chongqing Graduate Student Scientific Research Innovation Project under Grant CYB19174, in part by the Doctoral Innovative Talents Project of the Chongqing University of Posts and Telecommunications under Grant BYJS201810, in part by the National Natural Science Foundation of China under Grant 61972060, Grant U1713213, Grant 62027827, and Grant 41871226, in part by the Natural Science Foundation of Chongqing under Grant cstc2020jcyj-zdxmX0025 and Grant cstc2019cxcyljrc-td0270, and in part by the National Key Research, and Development Program of China under Grant 2019YFE0110800. *(Corresponding author: Weisheng Li.)*

Yidong Peng and Weisheng Li are with the Chongqing Key Laboratory of Image Cognition, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: eadonlife@yahoo.com; liws@cqupt.edu.cn).

Xiaobo Luo is with the Chongqing Engineering Research Center for Spatial Big Data Intelligent Technology, Chongqing Institute of Meteorological Science, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: luoxb@cqupt.edu.cn).

Jiao Du is with the School of Computer Science, and Cyber Engineering, Guangzhou University, Guangzhou 510006, China (e-mail: dujiao19880429@126.com).

Digital Object Identifier 10.1109/JSTARS.2021.3076170

I. Introduction

HYPERSPECTRAL imaging is an emerging modality that can provide the same scene over several hundreds of contiguous and narrow spectral bands. Owning to the wealth of available spectral information, hyperspectral images (HSIs) enhance our capabilities in many earth remote sensing tasks, including the monitoring and management of natural resources, the ecosystem, and disasters [1]. However, due to the inevitable trade-off between the spatial resolution, spectral resolution, and signal-to-noise ratio (SNR), HS imaging is generally limited by the poor spatial resolution of the data acquisition equipment [1]–[5]. Fortunately, the existing panchromatic and multispectral imaging cameras can provide panchromatic images (PANs) and multispectral images (MSIs) with much higher spatial resolution, which can be fused with low spatial resolution HSIs (LR-HSIs) to obtain HSIs with high spatial resolution (HR-HSIs). This procedure is often referred to as HSI superresolution, and it can be divided into two categories: hypersharpening and MS/HS fusion.

The earliest work on hypersharpening is an extension of Pansharpening [1], [6]–[8]. Pansharpening is a kind of fusion method that generates an HR-MSI by fusing an LR-MSI with a PAN. To date, a variety of pansharpening methods that take full advantage of the available multicomplementary spatial and spectral information in MSIs and PANs have been proposed [9]–[12]. These methods have been classified based on different rules [13]–[15]. More recently, Meng *et al.* [14] classified the existing pansharpening methods into component substitution, multiresolution analysis, and variational optimization. With the increasing availability of HS imaging systems, pansharpening was then extended to HSI superresolution by fusing HSIs with PANs, which was called hypersharpening [13], [16]–[19]. In addition, several of the hypersharpening methods also evolved from MS/HS fusion methods [4], [20]–[22]. In this case, the MSI is only composed of a single band, and thus, it can be reduced to a PAN image [13]. A more detailed comparison of hypersharpening methods can be found in [13].

In addition to hypersharpening methods, MS/HS fusion is another essential HSI superresolution category to obtain HR-HSIs. Representative MS/HS fusion approaches include matrix factorization-based methods [4], [20], [22]–[28], tensor-based methods [2], [3], [29]–[32], and deep learning-based methods [11], [33]–[36]. Deep learning has attracted increasing

attention in HSI superresolution due to its strong learning performance and high speed [29], [30], [34]. Li *et al.* [33] presented a deep spectral difference convolutional neural network model with a spatial constraint strategy for HSI superresolution. In [34], an HSI superresolution method based on a deep residual learning network that can learn the spectral prior of HSIs is proposed. Qu *et al.* [35] presented an unsupervised deep convolutional neural network using an unsupervised encoder–decoder architecture for HSI superresolution. However, deep learning-based methods usually require large amounts of samples for training the neural network to obtain the parameters of the network.

Matrix factorization-based methods have attracted increasing interest in recent years, and they have achieved promising performance. Their basic idea is that the unfolding matrix of an HR-HSI can be represented as a linear combination of a small number of distinct spectral signatures based on the assumption that the HR-HSI only contains a small amount of pure spectral signatures, namely, that the unfolding matrix of the HR-HSI can be factorized as a basis matrix (dictionary) multiplied by a coefficient matrix. Berné *et al.* [27] proposed a nonnegative matrix factorization (NMF) based method for mid-infrared astronomy. In [20], coupled nonnegative matrix factorization (CNMF) was proposed for HSI superresolution. In CNMF, the endmember and abundance matrices are estimated by alternating spectral unmixing based on NMF under the constraints of an observation model. However, the results from CNMF may not be always satisfactory since the NMF is often nonunique [26], [37]. Kawakami *et al.* [23] presented an HSI superresolution method based on a sparse matrix factorization technique, which can decompose an HR-HSI into a basis matrix and a sparse coefficient matrix. The HR-HSI can be then reconstructed using the spectral basis obtained from the LR-HSI and the sparse coefficient matrix estimated from the HR-MSI. Recently, for better inferences, the works in [24], [25], and[28] solve the superresolution problem by simultaneously exploiting the sparsity and nonnegativity constraints of the HR-HSI. For example, Wycoff *et al.* [28] presented a nonnegative sparse matrix factorization method to exploit both the sparsity and nonnegativity constraints of HR-HSIs. The superresolution problem was formulated as a joint optimization problem involving the nonnegative basis obtained from the LR-HSI and sparse coefficients estimated from the HR-MSI. In addition, the spatial similarities of the HR-HSI were also exploited to solve the superresolution problem [22], [26]. In [26], a nonnegative dictionary learning algorithm using the block-coordinate descent optimization technique is proposed to learn the spectral basis, and a structured sparse coding approach is used to estimate the coefficient matrix. In this way, the nonlocal spatial similarities of HR-HSIs are exploited to solve the superresolution problem and achieve promising performance.

Although many the matrix factorization-based methods have been proposed under different constraints and have yielded promising performances, all of them needed to unfold 3-D data structures into matrices, which makes it difficult to fully exploit the spatial-spectral correlations of HSIs [2], [3]. Since tensor-based techniques have a more powerful ability to simultaneously capture the correlations between and within different dimensions, they may be able to better deal with the HSI superresolution problem involving high-dimensional data, e.g., HSIs and MSIs, from the viewpoint of tensors [2], [3], [12]. Recently, the superresolution problem has been formulated from the viewpoint of tensor-based techniques [2], [3], [12], [29], [30]. Dian *et al.* [3] proposed a nonlocal sparse Tucker factorization. Although this method has promising performance, it failed to use the LR-HSI to estimate the core tensor. To address this problem, Li *et al.* [2] proposed a superresolution approach based on coupled sparse tensor factorization by formulating the estimation of the core tensor and dictionaries as a coupled sparse tensor factorization of the LR-HSI and HR-MSI. Zhang *et al.* [29] proposed a spatial-spectral-graph-regularized low-rank tensor decomposition-based HSI superresolution method by inferring the spatial smoothness and spectral consistency from the HR-MSI and LR-HSI, respectively. Dian *et al.* [30] proposed a low tensor train rank-based HSI superresolution method, which learns the correlations among the spatial, spectral, and nonlocal modes of the nonlocal similar HR-HSI cubes via a tensor train rank prior. Xu *et al.* [31] proposed a tensor-tensor product-based tensor sparse representation to model nonlocal patch tensors.

Most current tensor-based methods depend on two assumptions: 1) the LR-HSI and HR-MSI can be modeled as spectrally and spatially downsampled versions of the HR-HSI, respectively; and 2) the LR-HSI and HR-MSI can provide the high spatial-resolution information and high spectral-resolution information to the HR-HSI, respectively. However, this kind of methods sometimes suffers from spectral distortion after HSI superresolution because of the first unrealistic assumption. Moreover, these methods often use the ℓ_F-norm to describe the relations of these three datasets [2], [30], [31] or switch to the ℓ_1-norm when sparsity is desired [29], which indicates that those methods require substantial computational resources.

In this article, we propose a global gradient sparse and nonlocal low-rank tensor decomposition model with a hyper-Laplacian prior, which we refer to as HL-GSNLTD, for HSI superresolution by exploiting both the desired nonlocal and global properties of the HR-HSI. The framework of the proposed method is shown in Fig. 1. First, we investigated the error distribution of the upsampled LR-HSI and the target HR-HSI on the gradient domain and found that the gradients of the errors obey a hyper-Laplacian distribution (see Fig. 2). Hence, we would prefer to choose the hyper-Laplacian term depicted by the ℓ_p $(0 < p < 1)$-norm to describe the relationship of these two datasets. Second, to recover the spatial details, we adopt the assumption that the HR-MSI is modeled as a spatially downsampled version of the HR-HSI, which usually can be given in tensor-matrix multiplication format [2], [30], [31]. For an HR-HSI, spectrally adjacent bands are usually very similar, and the spatially adjacent pixels are highly correlated, which both indicate the fact that the HR-HSI has low-rank structures in both the spatial and spectral domains [29], [38]. Moreover, there are many similar cubes across the HR-HSI, and there exists strong correlation among those nonlocal similar cubes, which reveal that nonlocal similarity exists in the HR-HSI [30], [31]. To better exploit the nonlocal similarity and spatial-spectral low-rank structure of the HR-HSI, a new nonlocal low rank Tucker decomposition model is built to learn both the spatial

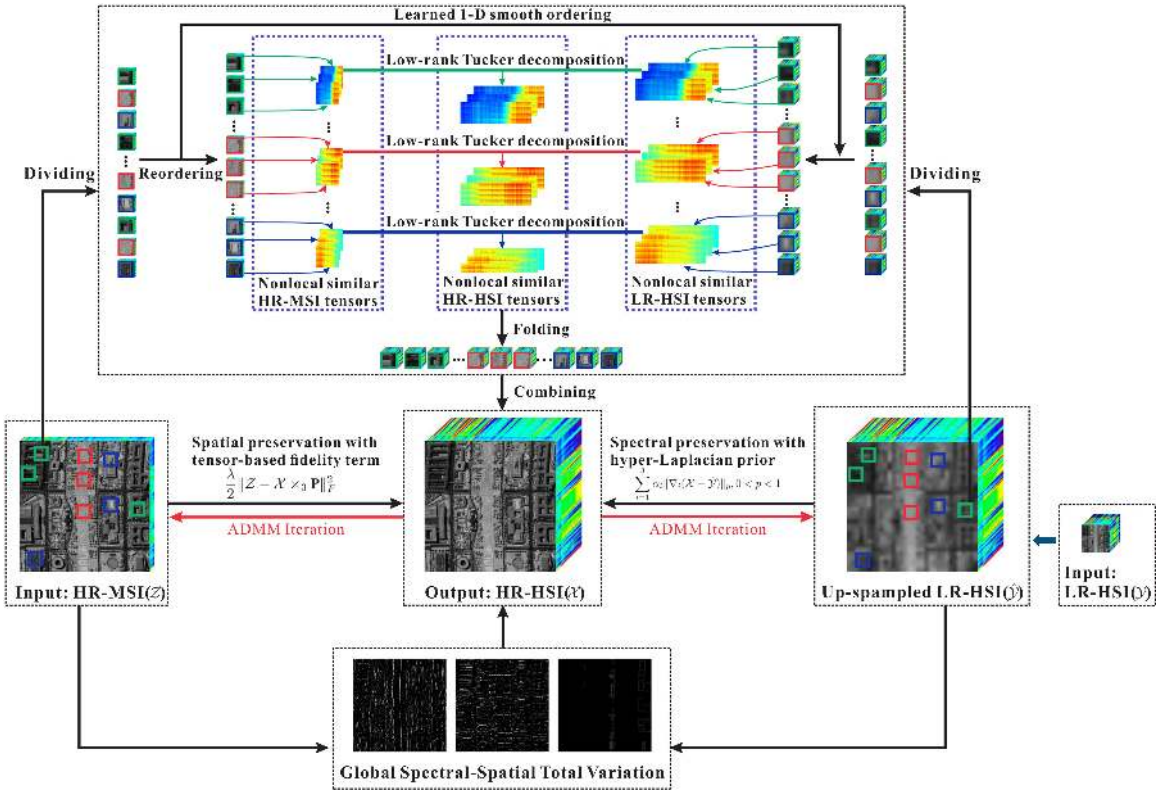

Fig. 1. Framework of the proposed HSI superresolution method.

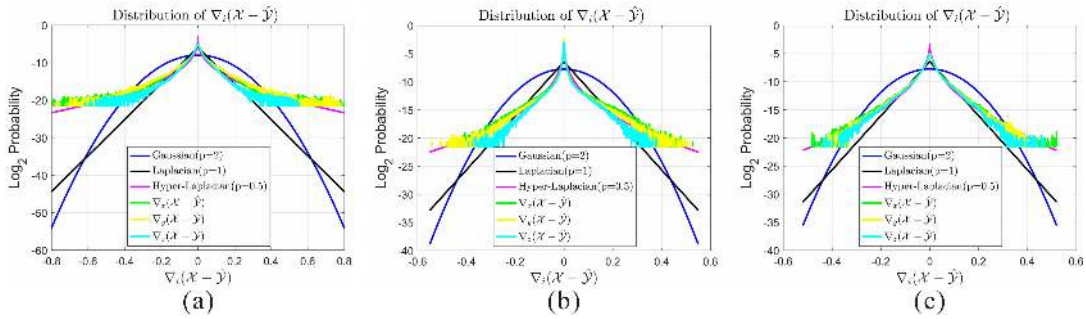

Fig. 2. Distributions of image gradient data $\nabla_i(\mathcal{X}-\hat{\mathcal{Y}})$ $(i=x,y,z)$, where the data, respectively, come from ROSIS-3, (a) HYDICE, (b) Hyperspec-VNIR-C, and (c) sensors.

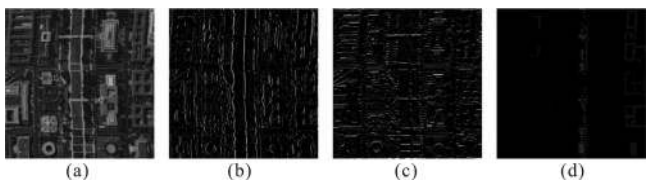

Fig. 3. Gradient images of the 3-D HR-HSI tensor along three different directions. (a) Original image. (b) Gradient image along the spatial horizontal direction. (c) Gradient image along the spatial vertical direction. (d) Gradient image along the spectral direction.

and spectral similarities of nonlocal cubes. However, this procedure processes all the grouped nonlocal similar cube tensors individually, which results in ignoring the global smoothness of the HR-HSI [38]–[42]. Fig. 3 depicts both the global spatial piecewise smooth property and spectral consistency of the target HR-HSI tensor along the spatial and spectral directions, respectively. Inspiringly, we also adopt a global spatial-spectral total-variation (SSTV)-regularization to reconstruct the HR-HSI from these nonlocal low-rank cubes, to further capture the global spatial piecewise smooth and spectral consistent structures of the HR-HSI. In this way, both the desired nonlocal and global properties of the HR-HSI are integrated into the proposed model.

In summary, the main contributions of this article are the following two aspects.

- According to image statistics, the gradients of the difference images of the upsampled LR-HSI and the target HR-HSI obey a hyper-Laplacian distribution, which correspond to the ℓ_p $(0<p<1)$-norm (see Fig. 2). Hence, the

ℓ_p term on the gradients of the difference images is more appropriate for enforcing spectral preservation compared with the ℓ_1 or ℓ_2 terms. Moreover, the ℓ_p term can reduce the complexity of the model without the blur kernel, which is generally assumed to be an inaccurate Gaussian one with an empirical variance. Under the framework of the alternating direction method of multipliers (ADMM) [43], the formulated nonconvex ℓ_p subproblem can be efficiently and readily solved by the generalized shrinkage/thresholding (GST) algorithm [44].

- A new global gradient sparse and nonlocal low-rank tensor decomposition model is introduced to characterize both the desired nonlocal and global properties of the HR-HSI. Specifically, the HR-HSI cubes are grouped into several 4-D tensors based on the learned 1-D smooth ordering calculated in the HR-MSI. Then, a new 3-D tensor is constructed by stacking each cube's transpose of the mode-3 unfolding matrix in the *k*th 4-D tensor. After the grouping operation, three modes of the 3-D tensor are highly correlated. Hence, these 3-D tensors can be naturally modeled as the form of low-rank tensor decomposition. Here, we apply a new low-rank Tucker decomposition model on these 3-D tensors to effectively capture the spatial and spectral similarities of nonlocal cubes. Moreover, global SSTV regularization is integrated into the fusion model to reconstruct the HR-HSI from these nonlocal low-rank cubes, to further capture the global spatial piecewise smoothness and spectral consistency of the HR-HSI.

II. Tensor Notations

A tensor is a multidimensional data array, whose order is the number of ways, also known as modes or dimensions. Following [45], lowercase letters denote vectors, e.g., $\boldsymbol{g}$, boldface capital letters denote matrices, e.g., $\mathbf{G}$, and calligraphic letters denote tensors, e.g., $\mathcal{G}$. Then, an *N*-dimensional tensor is defined as $\mathcal{G} \in \mathbb{R}^{I_1 \times \ldots \times I_n \times \ldots \times I_N}$, and $g_{i_1 \ldots i_2 \ldots i_N}$ represents its $(i_1, i_2 \ldots, i_N)$th entry. A fiber of tensor is a 1-D section defined by fixing all indices but one. The mode-*n* unfolding matrix, e.g., $\mathbf{G}_{(n)} = \text{Unfold}_n(\mathcal{G}) \in \mathbb{R}^{I_n \times I_1 I_2 \ldots I_{n-1} I_{n+1} \ldots I_N}$, is defined by arranging the mode-*n* fibers of the tensor as the columns of the matrix. Conversely, the inverse transform of unfolding can be defined as $\mathcal{G} = \text{fold}_n(\mathbf{G}_{(n)})$. The ℓ_1 norm of a tensor $\mathcal{G}$ is defined as $\|\mathcal{G}\|_1 = \sum_{i_1,\ldots,i_N} |g_{i_1,\ldots,i_N}|$, and the Frobenius norm is defined as $\|\mathcal{G}\|_F = \sqrt{\sum_{i_1,\ldots,i_N} |g_{i_1,\ldots,i_N}|^2}$.

The *n*-mode product of a tensor $\mathcal{G} \in \mathbb{R}^{I_1 \times \ldots \times I_n \times \ldots \times I_N}$ with a matrix $\mathbf{U} \in \mathbb{R}^{J \times I_n}$, denoted by $\mathcal{G} \times_n \mathbf{U}$, is also a tensor $\mathcal{A}$ with a size of $I_1 \times \cdots \times I_{n-1} \times J \times I_{n+1} \times \cdots \times I_N$. The *n*-mode product can also be expressed as each mode-*n* fiber multiplied by the matrix, namely, $\mathbf{A}_{(n)} = \mathbf{U}\mathbf{G}_{(n)}$.

Tucker decomposition [45], also known as a form of higher order principal component analysis, factors a tensor into a core tensor multiplied by the factor matrices along each mode as follows:

$$\mathcal{G} = \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times \cdots \times_N \mathbf{U}_N \tag{1}$$

where $\mathbf{U}_i \in \mathbb{R}^{I_i \times r_i}$ represents the factor matrix along the *i*th mode. $\mathcal{C} \in \mathbb{R}^{r_1 \times r_2 \times \cdots \times r_N}$ represents the core tensor that describes the level of interaction between the different factor matrices. Matricized form of Tucker decomposition can be defined as follows [45]:

$$\mathbf{G}_{(i)} = \mathbf{U}_i \mathbf{C}_{(i)} (\mathbf{U}_N \otimes \cdots \otimes \mathbf{U}_{i+1} \otimes \mathbf{U}_{i-1} \otimes \cdots \otimes \mathbf{U}_1)^{\mathsf{T}} \tag{2}$$

where $\otimes$ is the Kronecker product.

III. Proposed Model

In this article, the target HR-HSI is considered to be a tensor of order three, e.g., $\mathcal{X} \in \mathbb{R}^{N_w \times N_h \times N_s}$, where N_w and N_h are the height and width of the HR-HSI, respectively. N_s is the number of bands in the HR-HSI. The observed LR-HSI, which contains abundant high spectral-resolution information, is denoted by $\mathcal{Y} \in \mathbb{R}^{N_w^d \times N_h^d \times N_s}$, where $N_w^d = \frac{1}{r} \times N_w$ and $N_h^d = \frac{1}{r} \times N_h$. r is the spatial resolution ratio between the LR-HSI and HR-HSI. In addition, the tensor $\mathcal{Z} \in \mathbb{R}^{N_w \times N_h \times N_s^d}$ denotes the observed HR-MSI, which is a spectrally downsampled version of the target HR-HSI, i.e., $N_s^d < N_s$. The goal of superresolution is obtaining a reasonable HR-HSI $\mathcal{X}$ by utilizing the complementary information of the observations $\mathcal{Y}$ and $\mathcal{Z}$.

A. Spectral Preservation

For most of the image fusion models, the observed LR-HSI is considered to be a blurred and downsampled version of the target HR-HSI $\mathcal{X}$ [2], [30], [46]–[50]. However, the blur kernel is generally unknown, and it usually is assumed to follow a Gaussian distribution with an empirical variance, which actually is not fully reasonable. To get rid of the empirical blur kernel, we explore the relationship between the target HR-HSI and observed LR-HSI by investigating the gradient distribution of the difference image between these two datasets. Specifically, we interpolate the LR-HSI to obtain an upsampled LR-HSI $\hat{\mathcal{Y}}$, which has the same size as the target HR-HSI. Then, we investigated the gradient distributions of the difference images between the target HR-HSI and the upsampled LR-HSI on several datasets collected from different sensors (see Fig. 2.), and found that the gradient distributions along the horizontal, vertical and spectral directions all obey the hyper-Laplacian case (ℓ_p $(0 < p < 1)$) instead of a Gaussian (ℓ_2) or Laplacian (ℓ_1) distribution. Hence, it is more reasonable to apply an ℓ_p norm on the gradients of the difference images to describe the relationship between the target HR-HSI and the upsampled LR-HSI using *maximum a posteriori* theory. Hence, we define the energy term of the spectral preservation to be a nonconvex ℓ_p term as follows:

$$\mathbf{E}_{(1)} = \sum_{i=1}^{3} \alpha_i \|\nabla_i (\mathcal{X} - \hat{\mathcal{Y}})\|_p, \ 0 < p < 1 \tag{3}$$

where $\mathbf{E}_{(1)}$ is the energy term of the spectral preservation. ∇_i, $i = 1, 2, 3$ are the gradient operators along the horizontal, vertical, and spectral directions, respectively. α_i are positive parameters. $\|\cdot\|_p$ is the ℓ_p norm. From Fig. 2, we find that the hyper-Laplacian with $p = 1/2$ can fit these the gradients

very well; therefore, we believe that the $\ell_{1/2}$ norm is a more reasonable choice for depicting the relationship of the target HR-HSI and the upsampled LR-HSI on the gradient domain.

B. Spatial Structural Preservation

The observed HR-MSI $\mathcal{Z}$ contains abundant high spatial-resolution information, which can be used to provide the spatial details for the HR-HSI. In this article, we assume that the HR-MSI $\mathcal{Z}$ is modeled as a spectrally downsampled version of the HR-HSI $\mathcal{X}$. Thus, the energy term of the spatial structural preservation can be defined as follows:

$$\mathbf{E}_{(2)} = \frac{\lambda}{2}\|\mathcal{Z} - \mathcal{X} \times_3 \mathbf{P}\|_F^2 \tag{4}$$

where $\mathbf{P} \in \mathbb{R}^{N_s^d \times N_s}$ is the spectral response matrix. λ is a tradeoff parameter.

C. Nonlocal Low-Rank and Global Total Variation Assumptions

1) Nonlocal Low-Rank Assumption: The nonlocal similarity of an HR-HSI is a patch-based prior that describes the fact that a 3-D cube in the HSI has many similar cubes in the nearby space. Here, we present a nonlocal low-rank Tucker decomposition prior for the HR-HSI to capture the spatial and spectral similarities of nonlocal cubes. Since the spatial details of the latent HR-HSI mainly exist in the HR-MSI, the HR-MSI $\mathcal{Z}$ is first spatially segmented into overlapping 3-D cubes with the size of $d_w \times d_h \times N_s^d$, and then, the similar HR-MSI cubes are grouped into K clusters $\mathbf{Z}^{(k)} = \{\mathcal{Z}^{(k,j)} \in \mathbb{R}^{d_w \times d_h \times N_s^d}\}_{j=1}^{N_k}$, $k = 1, 2, \ldots, K$, where N_k is the number of cubes in the kth cluster, and (k, j) is the spatial location of the cube. Conventionally, there are some unsupervised clustering methods, such as k-means and k-means ++, can be employed in the clustering process [30], [51], [52]. However, the number of cluster centers must be set manually and the clustering result is highly dependent on the initialization of the cluster centers [31]. In this article, we use an image processing scheme based on the reordering of its patches to group the similar HR-MSI cubes [53]. Specifically, the segmented overlapping 3-D cubes are reordered to make sure that these segmented cubes are chained in the "shortest possible path," which can essentially solve the traveling salesman problem [53]. Then, we obtain a new ordering of the HR-MSI cubes that is expected to induce a highly regular (smooth or at least piecewise smooth) 1-D ordering, and the neighboring cubes in the new 1-D ordering are very similar and highly correlated. Finally, nonlocal similar cube sets can be constructed by extracting a limited number of consecutive cubes according to the learned 1-D smooth ordering.

Based on the learned 1-D smooth ordering computed in the HR-MSI, we construct clusters of HR-HSI cubes with the same spatial structure, i.e., $\mathbf{X}^{(k)} = \{\mathcal{X}^{(k,j)} \in \mathbb{R}^{d_w \times d_h \times N_s}\}_{j=1}^{N_k}$, where $\mathcal{X}^{(k,j)}$ has the same spatial locations as $\mathcal{Z}^{(k,j)}$. We stack the transpose of the mode-3 unfolding matrix of each cube in the *k*th cluster, and then we can obtain a new 3-D nonlocal similar cube tensor (NSCT), which is denoted as $\mathcal{X}_k \in \mathbb{R}^{d_w d_h \times N_s \times N_k}$, whose three modes, i.e., one spatial mode, one spectral mode, and one nonlocal mode, are highly correlated. Here, we define an operator $\mathcal{R}_k$: $\mathcal{X} \to \mathcal{X}_k$, where $\mathcal{R}_k$ is a binary operator that extracts the *k*th NSCT from a given tensor, i.e., $\mathcal{R}_k\mathcal{X} = \mathcal{X}_k$. For an NSCT that is highly regular, it can be effectively characterized using a form of low-rank tensor decomposition. Here, the low-rank constrained Tucker decomposition model is employed, and the nonlocal low-rank term is given as follows:

$$\begin{aligned}\mathbf{E}_{(3)}^{\mathbf{NL}} = &\sum_{k=1}^{K}\sum_{i=1}^{3}\beta_i\|\mathbf{V}_{ik}\|_* \\ &+ \frac{\mu}{2}\sum_{k=1}^{K}\|\mathcal{R}_k\mathcal{X} - \mathcal{H}_k \times_1 \mathbf{V}_{1k} \times_2 \mathbf{V}_{2k} \times_3 \mathbf{V}_{3k}\|_F^2\end{aligned} \tag{5}$$

where $\|\cdot\|_*$ is the nuclear norm, which is the sum of all singular values. $\mathcal{H}_k \in \mathbb{R}^{N_{J1} \times N_{J2} \times N_{J3}}$ is the core tensor of the *k*th NSCT. $\mathbf{V}_{ik}$, $i = 1, 2, 3$, are the factor matrices of the spatial, spectral, and nonlocal modes, respectively; and β_i is the weight with respect to (w.r.t.) factor matrix $\mathbf{V}_{ik}$. μ is a penalty parameter. It is noteworthy that the tensor rank is hard to determine. The work in [54] directly defined the tensor rank as the sum of the ranks of the unfolding matrix on all modes, which lacks a theoretical interpretation [29]; and the determination of the underlying rank of the unfolding matrix is difficult and unreliable [51]. In addition, the work in [2] empirically preseted the underlying rank of $\mathbf{V}_i$, which lacks versatility. In this article, we model the nonlocal low-rank property by imposing the nuclear norm on the factor matrices of the NSCTs.

2) Total Variation Assumption: Patch-based low-rank methods have achieved great success in HSI superresolution [29]–[31]. These patch-based methods partition the HR-HSI $\mathcal{X}$ into patches, fuse each patch cluster separately, and finally reconstruct the fused patches in a simple way, i.e., each pixel in the reconstructed HR-HSI is the weighted average across multiple patches containing the pixel. However, this procedure ignores the global spatial piecewise smoothness and spectral consistency of HR-HSI $\mathcal{X}$ (see Fig. 3). Here, we adopt the TV model to reconstruct the fused cubes. In addition, by considering that the target HR-HSI is the hyperspectral image but the multispectral image, whose two nearby band images are usually very similar, indicating spectral consistency. Therefore, we apply TV to both the spatial and spectral directions simultaneously to explore the global spatial piecewise smoothness and spectral consistency of the HR-HSI. The TV term for the latent HR-HSI is given as follows:

$$\mathbf{E}_{(3)}^{\mathbf{TV}} = \sum_{i=1}^{3} w_i\|\nabla_i\mathcal{X}\|_1 \tag{6}$$

which is a spatial-spectral TV (SSTV) regularizer, and w_i, $i = 1, 2, 3$ are the weights for the variations along two spatial directions and one spectral direction, respectively. $\|\cdot\|_1$ represents the ℓ_1 norm.

D. Proposed HL-GSNLTD Model

Combining the abovementioned four energy terms together, we finally obtain the following optimization problem:

$$\begin{aligned}\min_{\mathcal{X}} &\sum_{i=1}^{3}\alpha_i\|\nabla_i(\mathcal{X}-\hat{\mathcal{Y}})\|_{1/2}+\frac{\lambda}{2}\|\mathcal{Z}-\mathcal{X}\times_3\mathbf{P}\|_F^2\\&+\sum_{k=1}^{K}\sum_{i=1}^{3}\beta_i\|\mathbf{V}_{ik}\|_*+\sum_{i=1}^{3}w_i\|\nabla_i\mathcal{X}\|_1\\&+\frac{\mu}{2}\sum_{k=1}^{K}\|\mathcal{R}_k\mathcal{X}-\mathcal{H}_k\times_1\mathbf{V}_{1k}\times_2\mathbf{V}_{2k}\times_3\mathbf{V}_{3k}\|_F^2.\end{aligned}\tag{7}$$

It should be noted that, in the proposed HL-GSNLTD model, the nonlocal low-rank tensor decomposition is implemented on the NSCTs to better capture the spatial and spectral similarities of nonlocal cubes. On the other hand, the anisotropic SSTV regularizer is globally applied to reconstruct these NSCTs and effectively preserve both the spatial piecewise smoothness and spectral consistency of the whole HR-HSI. Using this strategy, the noise and outliers in the NSCTs can be efficiently dealt with, and the edge information of the latent HR-HSI can be preserved globally. It is clear that the proposed HL-GSNLTD model is a nonconvex and nonsmooth model that involves an $\ell_{1/2}$ term and an ℓ_1 term. In the following part, we will present the ADMM-based [43] algorithm for the proposed HL-GSNLTD model.

E. Optimization Algorithm

The optimization problem (7) can be solved by various methods. For efficiency, we use the ADMM framework [43] to solve it. First, by introducing three auxiliary variables $\mathcal{L}=\mathcal{X}$, $\mathcal{S}_i=\nabla_i(\mathcal{X}-\mathcal{Y})$, and $\mathcal{U}_i=\nabla_i\mathcal{X}$, where $i=1,2,3$, and a penalty term about $\mathcal{H}_k$ to prevent overfitting and improve stability, the augmented Lagrangian function can be exhibited as follows:

$$\begin{aligned}&L_u(\mathcal{S}_i,\mathcal{U}_i,\mathcal{L},\mathcal{X},\mathbf{V}_{ik},\mathcal{H}_k,\mathcal{A}_i,\mathcal{B},\mathcal{C}_i)\\&=\sum_{i=1}^{3}\alpha_i\|\mathcal{S}_i\|_{1/2}+\sum_{i=1}^{3}\frac{\gamma_i}{2}\|\mathcal{S}_i-\nabla_i(\mathcal{X}-\hat{\mathcal{Y}})+\mathcal{A}_i\|_F^2\\&+\frac{\lambda}{2}\|\mathcal{Z}-\mathcal{L}\times_3\mathbf{P}\|_F^2+\frac{\delta}{2}\|\mathcal{L}-\mathcal{X}+\mathcal{B}\|_F^2+\frac{\tau}{2}\sum_{k=1}^{K}\|\mathcal{H}_k\|_F^2\\&+\sum_{k=1}^{K}\sum_{i=1}^{3}\beta_i\|\mathbf{V}_{ik}\|_*+\frac{\mu}{2}\sum_{k=1}^{K}\|\mathcal{R}_k\mathcal{X}-\mathcal{H}_k\times_1\mathbf{V}_{1k}\times_2\mathbf{V}_{2k}\\&\times_3\mathbf{V}_{3k}\|_F^2+\sum_{i=1}^{3}w_i\|\mathcal{U}_i\|_1+\sum_{i=3}^{3}\frac{\eta_i}{2}\|\mathcal{U}_i-\nabla_i\mathcal{X}+\mathcal{C}_i\|_F^2\end{aligned}\tag{8}$$

where $\mathcal{A}_i$, $\mathcal{B}$, and $\mathcal{C}_i$ are the Lagrangian multipliers. δ, τ, η_i, and γ_i, where $i=1,2,3$, are the penalty parameters. The augmented Lagrangian function (8) can be minimized by iteratively optimizing the following subproblems.

1) Update $\mathbf{V}_{ik}$, $i=1,2,3$: After the mode-i unfolding operation, the sub-Lagrange function w.r.t. $\mathbf{V}_{ik}$ can be written as (9). Here, we take $\mathbf{V}_{1k}$ for instance to demonstrate the derivation process

$$\min_{\mathbf{V}_{1k}}\sum_{k=1}^{K}\beta_1\|\mathbf{V}_{1k}\|_*+\frac{\mu}{2}\sum_{k=1}^{K}\|\mathbf{V}_{1k}\mathbf{D}_{1k}-\mathbf{G}_{(1)}\|_F^2\tag{9}$$

where $\mathbf{G}_{(1)}$ is the mode-1 unfolding matrix of $\mathcal{G}=\mathcal{R}_k\mathcal{X}$, and $\mathbf{D}_{1k}=(\mathcal{H}_k\times_2\mathbf{V}_{2k}\times_3\mathbf{V}_{3k})_{(1)}$. However, it is quite difficult to optimize (9) due to the quadratic term of $\mathbf{V}_{1k}$. Recently, some proximal linearized algorithms for resolving such a problem have appeared [55]. Here, we first approximate the subproblem (9) by linearizing the quadratic term about $\mathbf{V}_{1k}$ so that the solution can be easily derived. With this linearization, the resulting approximation to (9) can be defined as

$$\min_{\mathbf{V}_{1k}}\sum_{k=1}^{K}\beta_1\|\mathbf{V}_{1k}\|_*+\frac{\mu}{2\theta}\sum_{k=1}^{K}\|\mathbf{V}_{1k}-(\mathbf{V}_{1k}^{\text{pre}}-\theta\mathbf{q}_{1k}^{pre})\|_F^2\tag{10}$$

where $\mathbf{V}_{1k}^{\text{pre}}$ is the estimated factor matrix in the previous iteration. $\mathbf{q}_{1k}^{\text{pre}}$ is the first-order derivative of (9) at $\mathbf{V}_{1k}$. $\theta>0$ is a proximal parameter. Then, this proximal problem can be efficiently solved using the singular value thresholding operator [56]. Finally, the solutions of the $\mathbf{V}_{2k}$- and $\mathbf{V}_{3k}$-related subproblems can be easily solved in the same way.

2) Update $\mathcal{H}_k$: According to (8), the $\mathcal{H}_k$-subproblem is presented as follows:

$$\begin{aligned}\min_{\mathcal{H}_k}\frac{\tau}{2}\sum_{k=1}^{K}\|\mathcal{H}_k\|_F^2+\frac{\mu}{2}\sum_{k=1}^{K}\|\mathcal{R}_k\mathcal{X}-\mathcal{H}_k\times_1\mathbf{V}_{1k}\\\times_2\mathbf{V}_{2k}\times_3\mathbf{V}_{3k}\|_F^2\end{aligned}\tag{11}$$

which can be transformed into matrix-vector form as

$$\min_{\boldsymbol{h}_k}\frac{\tau}{2}\sum_{k=1}^{K}\|\boldsymbol{h}_k\|_F^2+\frac{\mu}{2}\sum_{k=1}^{K}\|\boldsymbol{g}-\mathbf{Q}_k\boldsymbol{h}_k\|_F^2\tag{12}$$

where $\boldsymbol{h}_k=\text{vec}(\mathcal{H}_k)$ and $\boldsymbol{g}=\text{vec}(\mathcal{G})$ are obtained by vectorizing the tensors $\mathcal{H}_k$ and $\mathcal{G}$, respectively, where $\mathcal{G}=\mathcal{R}_k\mathcal{X}$; and $\text{vec}(\cdot)$ is the vectorization operation. $\mathbf{Q}_k=\mathbf{V}_{3k}\otimes\mathbf{V}_{2k}\otimes\mathbf{V}_{1k}$, where the symbol $\otimes$ represents the Kronecker product.

Problem (12) has a closed-form solution for each NSCT

$$\boldsymbol{h}_k=(\tau\mathbf{I}+\mu\mathbf{Q}_k^T\mathbf{Q}_k)^{-1}(\mu\mathbf{Q}_k^T\boldsymbol{g}).\tag{13}$$

It seems that (13) is hard to calculate because of the involved Kronecker product. However, the computation cost for each NSCT, in fact, is quite small because (13) calculates each small size NSCT separately. In additon, we also employ the following strategy [2] to alleviate this problem. We note that

$$\begin{aligned}(\tau\mathbf{I}+\mu\mathbf{Q}_k^T\mathbf{Q}_k)^{-1}=&(\Psi_3\otimes\Psi_2\otimes\Psi_1)\\&(\mu(\Omega_3\otimes\Omega_2\otimes\Omega_1)+\tau\mathbf{I})^{-1}\\&(\Psi_3^T\otimes\Psi_2^T\otimes\Psi_1^T)\end{aligned}\tag{14}$$

where Ψ_i and Ω_i, $i=1,2,3$, are unitary matrices and nonnegative diagonal matrices consisting of the eigenvectors and eigenvalues of $\mathbf{V}_{ik}^T\mathbf{V}_{ik}$, respectively. Therefore, $(\mu(\Omega_3\otimes\Omega_2\otimes\Omega_1)+\tau\mathbf{I})^{-1}$ is diagonal and easy to calculate. Moreover, the operations on Ψ_i and Ψ_i^T are just i-mode products, and the operation on $(\mu(\Omega_3\otimes\Omega_2\otimes\Omega_1)+\tau\mathbf{I})^{-1}$ is an elementwise

Algorithm 1: Generalized Shrinkage/Thresholding (GST) Algorithm.

Input: y, λ, p, J

$\tau_p^{GST}(\lambda) = (2\lambda(1-p))^{\frac{1}{2-p}} + \lambda p(2\lambda(1-p))^{\frac{p-1}{2-p}}$
if $|y| \le \tau_p^{GST}(\lambda)$ **then**
  $\mathbf{T}_p^{GST}(y;\lambda) = 0$
else
  $k = 0, x^{(k)} = |y|$
  for $k = 0, 1, \ldots, J$ **do**
    $x^{(k+1)} = |y| - \lambda p(x^{(k)})^{p-1}$
  end for
  $\mathbf{T}_p^{GST}(y;\lambda) = \mathrm{sgn}(y)x^k$
end if

Output: $\mathbf{T}_p^{GST}(y;\lambda)$

multiplication. Besides, the term $\mu\mathbf{Q}_k^T\boldsymbol{g}$ in (13) can be calculated as follows:

$$\mu\mathbf{Q}_k^T\boldsymbol{g} = \mu\mathrm{vec}(\mathcal{G} \times_1 \mathbf{V}_{1k} \times_2 \mathbf{V}_{2k} \times_3 \mathbf{V}_{3k}). \quad (15)$$

3) Update $\mathcal{S}_i$: By fixing the other variables, the optimization w.r.t. $\mathcal{S}_i$, where $i = 1, 2, 3$, can be equivalent to the following:

$$\min_{\mathcal{S}_i} \sum_{i=1}^{3}\alpha_i\|\mathcal{S}_i\|_{1/2} + \sum_{i=3}^{3}\frac{\gamma_i}{2}\|\mathcal{S}_i - \nabla_i(\mathcal{X} - \hat{\mathcal{Y}}) + \mathcal{A}_i\|_F^2 \quad (16)$$

where problem (16) is a nonconvex sparse coding problem that is calculated using the $\ell_{1/2}$ norm. This indicates that it is quite difficult to optimize (16) with guarantees of global convergence. Fortunately, a number of algorithms have been proposed for solving such a problem, e.g., the iteratively reweighted least squares [57], the iteratively thresholding method [58], [59], the look-up table [60], and the GST algorithm [44]. Here, the GST [44] algorithm (summarized in Algorithm 1) is adopted to solve the given nonconvex problem owning to its simple scheme and fast speed, and then problem (16) has the following solution:

$$\mathcal{S}_i = \mathbf{T}_{1/2}^{GST}(\nabla_i(\mathcal{X} - \hat{\mathcal{Y}}) - \mathcal{A}_i, \frac{\alpha_i}{\gamma_i}),\ i = 1, 2, 3. \quad (17)$$

For more details about the GST algorithm, please refer to [44].

4) Update $\mathcal{U}_i$: From (8), the subproblem w.r.t. $\mathcal{U}_i$, where $i = 1, 2, 3$, can be written as

$$\min_{\mathcal{U}_i} \sum_{i=1}^{3} w_i\|\mathcal{U}_i\|_1 + \sum_{i=3}^{3}\frac{\eta_i}{2}\|\mathcal{U}_i - \nabla_i\mathcal{X} + \mathcal{C}_i\|_F^2. \quad (18)$$

Problem (18) is strongly convex, and it can be efficiently solved using the soft-thresholding (shrinkage) operator [56]

$$\mathbf{Soft}(a, b) = \begin{cases} a - b, & \text{if } a > b \\ a + b, & \text{if } a < b \\ 0, & \text{otherwise.} \end{cases} \quad (19)$$

Then, we can update $\mathcal{U}_i$ as

$$\mathcal{U}_i = \mathbf{Soft}(\nabla_i\mathcal{X} - \mathcal{C}_i, \frac{w_i}{\eta_i}), i = 1, 2, 3. \quad (20)$$

5) Update $\mathcal{L}$: For the $\mathcal{L}$-subproblem, it has the following minimization problem:

$$\min_{\mathcal{L}} \frac{\lambda}{2}\|\mathcal{Z} - \mathcal{L} \times_3 \mathbf{P}\|_F^2 + \frac{\delta}{2}\|\mathcal{L} - \mathcal{X} + \mathcal{B}\|_F^2. \quad (21)$$

Using the properties of mode-n matrix unfolding, the subproblem becomes

$$\min_{\mathbf{L}_{(3)}} \frac{\lambda}{2}\|\mathbf{Z}_{(3)} - \mathbf{P}\mathbf{L}_{(3)}\|_F^2 + \frac{\delta}{2}\|\mathbf{L}_{(3)} - \mathbf{X}_{(3)} + \mathbf{B}_{(3)}\|_F^2 \quad (22)$$

where $\mathbf{L}_{(3)}$, $\mathbf{Z}_{(3)}$, $\mathbf{X}_{(3)}$, and $\mathbf{B}_{(3)}$ are the mode-3 unfolding matrices of tensors $\mathcal{L}$, $\mathcal{Z}$, $\mathcal{X}$, and $\mathcal{B}$, respectively. Problem (22) is strongly convex. Hence, minimizing (22) w.r.t. $\mathbf{L}_{(3)}$ is equivalent to forcing the derivative of (22) to be zero, leading to the following closed-form solution:

$$\mathbf{L}_{(3)} = (\lambda\mathbf{P}^T\mathbf{P} + \delta\mathbf{I})^{-1}(\lambda\mathbf{P}^T\mathbf{Z}_{(3)} + \delta(\mathbf{X}_{(3)} - \mathbf{B}_{(3)})) \quad (23)$$

where $\mathbf{I} \in \mathbb{R}^{N_s \times N_s}$ is the identity matrix. Since the solution of the $\mathbf{L}_{(3)}$-subproblem only involves an inverse computation of small matrices with a size of $N_s \times N_s$, its computation cost is quite small. After obtaining the mode-3 unfolding matrix $\mathbf{L}_{(3)}$, we can acquire the tensor $\mathcal{L}$ using $\mathcal{L} = \mathrm{fold}_3(\mathbf{L}_{(3)})$.

6) Update $\mathcal{X}$: From (8), we have

$$\begin{aligned}\min_{\mathcal{X}} &\sum_{i=1}^{3}\frac{\gamma_i}{2}\|\mathcal{S}_i - \nabla_i(\mathcal{X} - \hat{\mathcal{Y}}) + \mathcal{A}_i\|_F^2 + \frac{\delta}{2}\|\mathcal{L} - \mathcal{X} + \mathcal{B}\|_F^2 \\ &+ \frac{\mu}{2}\sum_{k=1}^{K}\|\mathcal{R}_k\mathcal{X} - \mathcal{H}_k \times_1 \mathbf{V}_{1k} \times_2 \mathbf{V}_{2k} \times_3 \mathbf{V}_{3k}\|_F^2 \quad (24) \\ &+ \sum_{i=3}^{3}\frac{\eta_i}{2}\|\mathcal{U}_i - \nabla_i\mathcal{X} + \mathcal{C}_i\|_F^2\end{aligned}$$

which can be efficiently solved by the fast Fourier transform (FFT) method under the periodic boundary condition

$$\mathcal{X} = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(\mathcal{W})}{\mathcal{F}(\mathcal{K})}\right) \quad (25)$$

where the $\mathcal{F}$ and $\mathcal{F}^{-1}$ are the FFT and inverse FFT, respectively, and

$$\begin{aligned}\mathcal{W} = &\sum_{i=1}^{3}\gamma_i\nabla_i^T(\nabla_i\hat{\mathcal{Y}} + \mathcal{S}_i + \mathcal{A}_i) + \delta(\mathcal{L} + \mathcal{B}) \\ &+ \mu\sum_{k=1}^{K}\mathcal{R}_k^T(\mathcal{H}_k \times_1 \mathbf{V}_{1k} \times_2 \mathbf{V}_{2k} \times_3 \mathbf{V}_{3k}) \quad (26) \\ &+ \sum_{i=3}^{3}\eta_i\nabla_i^T(\mathcal{U}_i + \mathcal{C}_i)\end{aligned}$$

and

$$\mathcal{K} = \sum_{i=1}^{3}(\gamma_i + \eta_i)\nabla_i^T\nabla_i + \delta\mathcal{O} + \mu\sum_{k=1}^{K}\mathcal{R}_k^T\mathcal{R}_k \quad (27)$$

where $\mathcal{O} \in \mathbb{R}^{N_w \times N_h \times N_s}$ is a tensor whose elements are 1.

Algorithm 2: HL-GSNLTD for HSI Super-Resolution.

Input: $\hat{\mathcal{Y}}$, $\mathcal{Z}$, $\mathbf{P}$, N_k, τ, λ, α_i, β_i, w_i, $i = 1, 2, 3$
Extracting nonlocal tensors
(a) Group HR-MSI cubes via the reordering-based image processing scheme [53]
(b) Group HR-HSI cubes using the learned 1-D smooth ordering calculated in HR-MSI
(c) Obtain new NSCT by stacking each cube's transpose of mode-3 unfolding matrix in HR-HSI cube cluster
Initialization: Initialize $\mathbf{V}_{ik}$, $i = 1, 2, 3$ via the VCA [61] algorithm, $\mathcal{H}_k = 0$, $\mathcal{S}_i = \mathcal{A}_i = \mathcal{C}_i = 0$, $\mathcal{L} = \mathcal{B} = 0$, $\gamma_i = \eta_i = 10^{-2} (i = 1, 2)$, $\delta = \gamma_3 = \eta_3 = 10^{-3}$, $\mu = 0.1$, $maxIter = 15$, and $iter = 0$
Solving (7) via ADMM
while *not converged and* $iter < maxIter$ **do**
  $iter = iter + 1$
  for *k=1:K* **do**
    Update $\mathbf{V}_{ik}^{iter}, i = 1, 2, 3$ by (10)
    Update $\mathbf{H}_k^{iter}$ by (13)
  end
  Update $\mathcal{S}_i^{iter}, i = 1, 2, 3$ via (16) with **Algorithm 1**
  Update $\mathcal{U}_i^{iter}, i = 1, 2, 3$ via (18) with soft-thresholding algorithm
  Update $\mathcal{L}^{iter}$ by (23)
  Update $\mathcal{X}^{iter}$ via (25) with the FFT method
  Update the multipliers $\mathcal{A}_i^{iter}, \mathcal{C}_i^{iter}, i = 1, 2, 3$, and $\mathcal{B}^{iter}$ by (28)
  Check the convergence condition: $\max\{\sum_{i=1}^{3} \|\mathcal{S}_i^{iter} - \nabla_i(\mathcal{X}^{iter} - \hat{\mathcal{Y}}^{iter})\|_F^2, \sum_{i=3}^{3} \|\mathcal{U}_i^{iter} - \nabla_i \mathcal{X}^{iter}\|_F^2, \|\mathcal{X}^{iter} - (\sum_{k=1}^{K} \mathcal{R}_k^T \mathcal{R}_k)^{-1} \sum_{k=1}^{K} \mathcal{R}_k^T (\mathcal{H}_k^{iter} \times_1 \mathbf{V}_{1k}^{iter} \times_2 \mathbf{V}_{2k}^{iter} \times_3 \mathbf{V}_{3k}^{iter})\|_F^2\} \leq \varepsilon$
end
Output: Fused image $\mathcal{X}$

7) Update the Multipliers: After solving each subproblem of problem (8), the multipliers $\mathcal{A}_i$, $\mathcal{B}$, and $\mathcal{C}_i$ can be updated by the following formulas:

$$\begin{aligned} \mathcal{A}_i &= \mathcal{A}_i + (\mathcal{S}_i - \nabla_i(\mathcal{X} - \hat{\mathcal{Y}})), i = 1, 2, 3 \\ \mathcal{B} &= \mathcal{B} + (\mathcal{L} - \mathcal{X}) \\ \mathcal{C}_i &= \mathcal{C}_i + (\mathcal{U}_i - \nabla_i \mathcal{X}), i = 1, 2, 3. \end{aligned} \tag{28}$$

Using the ADMM framework, the difficult minimization problem (8) is split into six simpler subproblems that are solved iteratively. Note that all these six subproblems can be effectively solved using fast and accurate techniques. For instance, the $\mathbf{V}_{ik}$ subproblem can be easily solved by the singular value thresholding operator after linearizing its quadratic term, and the nonconvex $\mathcal{S}_i$ subproblem can be readily solved by the fast GST algorithm. Moreover, the $\mathcal{U}_i$ and $\mathcal{X}$ subproblems can be efficiently solved by the soft-thresholding strategy and the FFT algorithm, respectively. Specifically, the complete algorithm of the proposed method (8) is summarized in Algorithm 2.

IV. Experimental Study

In this section, we conduct an extensive series of image fusion experiments on both the synthetic and real data sets to verify the effectiveness of the proposed method. First, we introduce the synthetic datasets and the evaluation indices used in the experiments. Second, we present the experimental results on the synthetic datasets and the real dataset. Then, the influences of the parameters and energy terms, and the computational time are analyzed.

A. Synthetic Datasets

1) ROSIS-3 University of Pavia: The first dataset was taken by the reflective optics spectrographic imaging system (ROSIS-3) optical airborne sensor over the University of Pavia, Italy, in 2003. The image size is $610 \times 340 \times 115$ with a ground sampling distance (GSD) of 1.3 m and spectral coverage from 0.43 to 0.86 μm. After removing the water vapor absorption bands, a total of 103 bands with a spectral coverage from 0.43 to 0.838 μm remained. In our experiments, we select the middle-left $256 \times 256 \times 93$-pixel-size as the reference image.

2) HYDICE Washington DC Mall: The second dataset is Washington DC Mall data acquired by the hyperspectral digital imagery collection experiment (HYDICE) sensor over the Mall in Washington, DC, USA, in 1995. The image size is $1280 \times 307 \times 210$ with a GSD of 2.5 m and spectral coverage from 0.4 to 2.5 μm. The number of bands is reduced to 191 spectral bands with a spectral coverage from 0.9 to 1.4 μm after removing the water vapor absorption bands. In the experiments, a $256 \times 256 \times 93$ cube size image is cropped as the reference image.

3) Hyperspec-VNIR Chikusei: The third dataset was acquired by the Headwalls Hyperspec-VNIR-C imaging sensor over Chikusei, Ibaraki, Japan, in 2014. It consists of 2517×2335 pixel images with a GSD of 2.5 m and 128 spectral bands with a spectral coverage ranging from 0.363 to 1.018 μm. The scene is mainly agricultural and urban areas. We select the down-left $256 \times 256 \times 128$-pixel-size as the reference image.

B. Implementation Details

In the experiments, the fusing performance is influenced by the cube size and number of cubes for the nonlocal similarity search and the parameters for the optimization procedure. The cube size $d_w \times d_h \times N_s$ for the nonlocal similarity search is set to $4 \times 4 \times N_s$ with step size of 3, and the number of cubes in each NSCT was set to $N_k = 6$. Then, the size of the NSCT is determined as $16 \times N_s \times 6$. To reduce the influence of the initialization strategy, we carefully initialize the factor matrices using the fast vertex component analysis (VCA) algorithm [61]. Considering that the HR-MSI contains important high spatial resolution information and the spectral information mostly exists in the LR-HSI, we, hence, initialize $\mathbf{V}_{1k} \in \mathbb{R}^{d_w * d_h \times N_{J1}}$ and $\mathbf{V}_{3k} \in \mathbb{R}^{N_k \times N_{J3}}$ using the mode-1 (spatial mode) and mode-3 (nonlocal mode) unfolding matrices of the NSCTs extracted from the HR-MSI, respectively, where N_{J1} and N_{J3} are the number of atoms of $\mathbf{V}_{1k}$ and $\mathbf{V}_{3k}$, respectively. $\mathbf{V}_{2k} \in \mathbb{R}^{N_s \times N_{J2}}$

is initialized using the mode-2 (spectral mode) unfolding matrix of the NSCTs extracted from the upsampled LR-HSI, where N_{J2} is the number of atoms of $\mathbf{V}_{2k}$. As we know that it is not easy to estimate the mode-k rank for the desired $\mathcal{X}_k$ directly, we, hence, directly set N_{J1}, N_{J2}, and N_{J3} as $d_w * d_h$, N_s, and N_k, respectively, and then the low-rank constraints are achieved by the weights β_i alternatively. The weights for each mode are set as

$$\begin{aligned} &\beta_i = \rho\sqrt{N_{\max}/N_{Ji}}, N_{\max} = \max(N_{J1}, N_{J2}, N_{J3}) \\ &\tau = 1/\|\hat{\mathcal{Y}}\|_F^2 \end{aligned} \tag{29}$$

where ρ is a ratio parameter set to 0.01. Considering the balance of the terms in (8), the regularization parameters should be carefully designed in the proposed model. In the experiments, we empirically set $\alpha_1 = \alpha_2 = 1 \times 10^{-3}, \alpha_3 = 5 \times 10^{-2}, w_1 = w_2 = 1 \times 10^{-2}, w_3 = 5 \times 10^{-2}, \mu = 1 \times 10^{-1}, \lambda = 5 \times 10^{-1}$, $\delta = \eta_3 = \gamma_3 = 1 \times 10^{-3}$, and $\eta_i = \gamma_i = 1 \times 10^{-2}$ $(i = 1, 2)$ for all datasets. Although the proposed method contains many parameters, most of them are relatively robust to different datasets. Detailed discussions of the selections of parameters will be presented in Section V-C.

In this article, we compare our results with those of five current state-of-the-art approaches, such as nonnegative structured sparse representation (NSSR) [26], coupled sparse tensor factorization (CSTF) [2], low tensor train rank (LTTR) based method [30], spatial-spectral-graph-regularized low-lank tensor decomposition (SSGLRTD) [29], and deep HSI sharpening (DHSIS) [34]. The parameters for the compared methods are set to the default values or determined according to the suggestions in their articles to obtain the best performance. For NSSR, the number of atoms of the dictionary is set to 55, the parameters η_1 and η_2 are set to 1.5×10^{-4} and 1×10^{-4}, respectively. As a deep learning method, DHSIS requires large numbers of samples to train the neural network. The University of Pavia and Washington DC Mall datasets, which do not have enough image samples to train a neural network, are unsuitable for DHSIS. Thus, DHSIS was only implemented on the Hyperspec Chikusei dataset. For the Hyperspec Chikusei dataset, we extract the training samples from the original images outside the test region in the experiments. The numbers of training samples are 20 614 for the Hyperspec Chikusei dataset.

In addition, five quantitative metrics are selected to quantitatively evaluate the quality of the reconstructed HSIs, such as the peak signal-noise-ratio (PSNR), spectral angle mapper (SAM) [62], degree of distortion (DD) [2], *Erreur Relative Globale Adimensionnelle de Synthèse* (ERGAS) [63], and universal image quality index (UIQI) [64]. The larger the PSNR and the smaller the SAM, DD, and ERGAS are, the better the fusion results. The UIQI has a range of [−1,1], and its ideal value is 1. The larger the UIQI value is, the better the fusion results.

C. Experimental Results on Synthetic Datasets

In this section, we use the original University of Pavia and Washington DC Mall images as the reference HR-HSIs $\mathcal{X}$. The LR-HSIs $\mathcal{Y}$ are simulated by uniformly averaging over disjoint $r \times r$ blocks of the reference HR-HSIs (e.g., $r = 4$, 8, 16). The HR-MSIs $\mathcal{Z}$ are generated by filtering the reference HR-HSIs using the spectral reflectance response files of the Red, Green, Blue (R.G.B.), and Near Infrared (NIR) bands of the IKONOS multispectral sensor. Tables I and II present the PSNR, SAM, DD, ERGAS, and UIQI values of the reconstructed HSIs for the University of Pavia and Washington DC Mall datasets, respectively. The best values are marked in bold for clarity. The first and third rows of Figs. 4 and 5 show the reconstructed images of the University of Pavia and Washington DC Mall datasets with the scaling factors 8 and 16. The reconstructed images of the University of Pavia dataset are displayed as the composite images produced by the 50th, 23th, and 5th band images. The reconstructed images of the Washington DC Mall dataset are displayed as the composite images produced by the 60th, 30th, and 8th band images. The pixelwise root-mean-square error (RMSE) [1] images are used to visualize the errors between the reconstructed results and reference image (second and fourth rows in Figs. 4 and 5), where blue pixels indicate no fusion error

TABLE I
QUANTITATIVE EVALUATION OF THE COMPARED METHODS ON THE UNIVERSITY OF PAVIA DATASET WITH UNIFORM BLUR

Method	NSSR	CSTF	SSGLRTD	LTTR	Proposed
$r = 4$					
PSNR	42.607	41.407	42.230	42.652	**43.819**
SAM	2.461	2.841	2.471	2.421	**2.076**
DD	1.520	1.728	1.643	1.469	**1.302**
ERGAS	1.254	1.267	1.260	1.232	**1.028**
UIQI	0.991	0.992	0.989	**0.993**	**0.993**
$r = 8$					
PSNR	41.187	40.324	40.869	41.663	**43.343**
SAM	2.604	3.027	2.636	2.527	**2.270**
DD	1.808	1.991	1.847	1.752	**1.366**
ERGAS	0.725	0.799	0.762	0.684	**0.592**
UIQI	0.987	0.991	0.989	0.989	**0.992**
$r = 16$					
PSNR	39.537	40.309	40.056	41.361	**42.621**
SAM	3.650	3.272	3.380	2.943	**2.507**
DD	2.358	2.011	2.063	1.818	**1.472**
ERGAS	0.470	0.397	0.408	0.359	**0.319**
UIQI	0.982	0.990	0.988	0.989	**0.991**

TABLE II
QUANTITATIVE EVALUATION OF THE COMPARED METHODS ON THE WASHINGTON DC MALL DATASET WITH UNIFORM BLUR

Method	NSSR	CSTF	SSGLRTD	LTTR	Proposed
$r = 4$					
PSNR	42.318	40.181	40.695	42.709	**42.775**
SAM	2.318	2.370	2.682	2.127	**2.118**
DD	1.716	2.064	2.031	1.601	**1.557**
ERGAS	1.201	1.324	1.591	1.179	**1.132**
UIQI	0.993	0.991	0.988	**0.995**	0.994
$r = 8$					
PSNR	40.594	39.471	39.131	41.738	**42.027**
SAM	2.819	2.617	2.871	2.534	**2.231**
DD	2.144	2.316	2.431	1.870	**1.697**
ERGAS	0.765	0.760	0.837	0.666	**0.627**
UIQI	0.990	0.990	0.987	0.991	**0.993**
$r = 16$					
PSNR	39.592	39.073	38.722	40.542	**41.623**
SAM	3.361	2.928	3.414	2.778	**2.451**
DD	2.520	2.601	2.662	2.186	**1.852**
ERGAS	0.416	0.446	0.456	0.402	**0.342**
UIQI	0.988	0.986	0.979	0.987	**0.992**

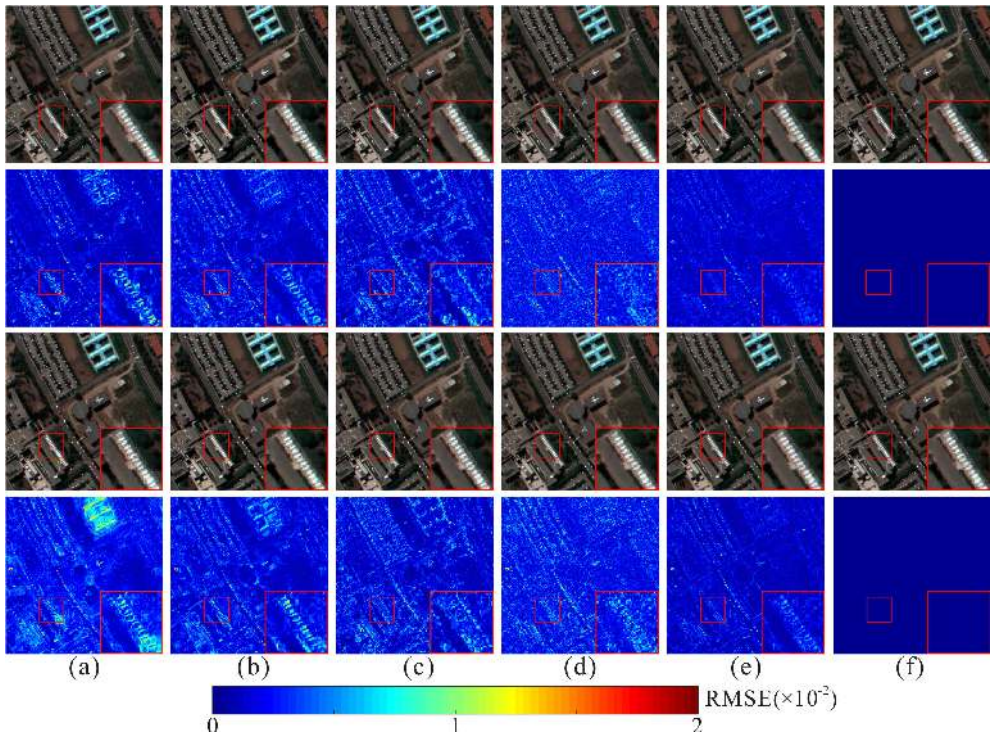

Fig. 4. Reconstructed images and the corresponding RMSE images of the University of Pavia dataset with uniform blur and scaling factor (Top two rows) $r = 8$ and (Bottom two rows) $r = 16$. (a) NSSR [26]. (b) CSTF [2]. (c) SSGLRTD [29]. (d) LTTR [30]. (e) Proposed method. (f) Reference image.

and red pixels mean severe local fusion errors. Besides, we magnify a meaningful region of each reconstructed image for better visual comparison. From magnified regions, the NSSR, CSTF, and SSGLRTD methods produce clear and sharp spatial details. However, there are many obvious artifacts in some areas of their RMSE images, indicating their poor performance in those areas. Although LTTR produces few artifacts in its RMSE images, there are many noises distributed over the entire RMSE images. Compared with other reconstructed results, the reconstructed image of the proposed method provides more consistent spectral information and preserves more spatial details. Fig. 6 shows the spectral curves of the reconstructed results with uniform blur and scaling factor $r = 8$ to compare the performances in terms of spectral preservation. From Fig. 6, we obtain that the spectral signatures of the proposed method are more consistent with the reference spectral signatures than the other compared methods for the entire spectral coverage. The proposed method yields better performance compared with the other four methods both quantitatively and qualitatively in most cases.

We also demonstrate the performance of the proposed method to the Gaussian blur on the University of Pavia and Washington DC Mall datasets. To generate the LR-HSIs, we first apply an 8×8 Gaussian blur function of standard deviation 2.15 to the reference HR-HSIs before downsampling along the width and height modes of the reference HR-HSIs with scaling factor 8. Table III presents the quantitative evaluation of the test methods on the University of Pavia dataset and Washington DC Mall dataset with Gaussian blur and scaling factor 8. From Table III, we obtain that the proposed method produces more competitive performance in the spatial and spectral quality metrics. This indicates that the proposed method performs better at preserving the spectral and spatial information than the other compared methods.

TABLE III
QUANTITATIVE EVALUATION OF THE COMPARED METHODS ON THE UNIVERSITY OF PAVIA DATASET AND WASHINGTON DC MALL DATASET WITH GAUSSIAN BLUR AND SCALING FACTOR 8

Method	NSSR	CSTF	SSGLRTD	LTTR	Proposed
Pavia data set					
PSNR	40.894	39.745	40.686	41.384	**43.033**
SAM	2.652	3.218	2.694	2.598	**2.315**
DD	1.852	1.231	1.882	1.813	**1.428**
ERGAS	0.769	0.844	0.789	0.691	**0.618**
UIQI	0.987	0.990	0.989	0.988	**0.992**
Washington data set					
PSNR	40.248	39.023	38.972	41.482	**41.899**
SAM	2.856	2.902	2.936	2.575	**2.293**
DD	2.228	2.419	2.478	1.934	**1.728**
ERGAS	0.781	0.829	0.858	0.684	**0.635**
UIQI	0.990	0.987	0.987	0.991	**0.992**

D. Experimental Results on Synthetic Dataset Corrupted by Gaussian Noise

The objective of this section is to verify the robustness of the proposed method to the noise on the Hyperspec Chikusei data with Gaussian noise. We produce the noisy LR-HSI $\mathcal{Y}$ and HR-MSI $\mathcal{Z}$ in the same way described in work [26]. Specifically, the noisy LR-HSI $\mathcal{Y}$ is produced by first applying an 8×8 Gaussian blur with standard deviation 3 before downsampling with scaling factor 8, and then adding the Gaussian white noise. Here, SNRh

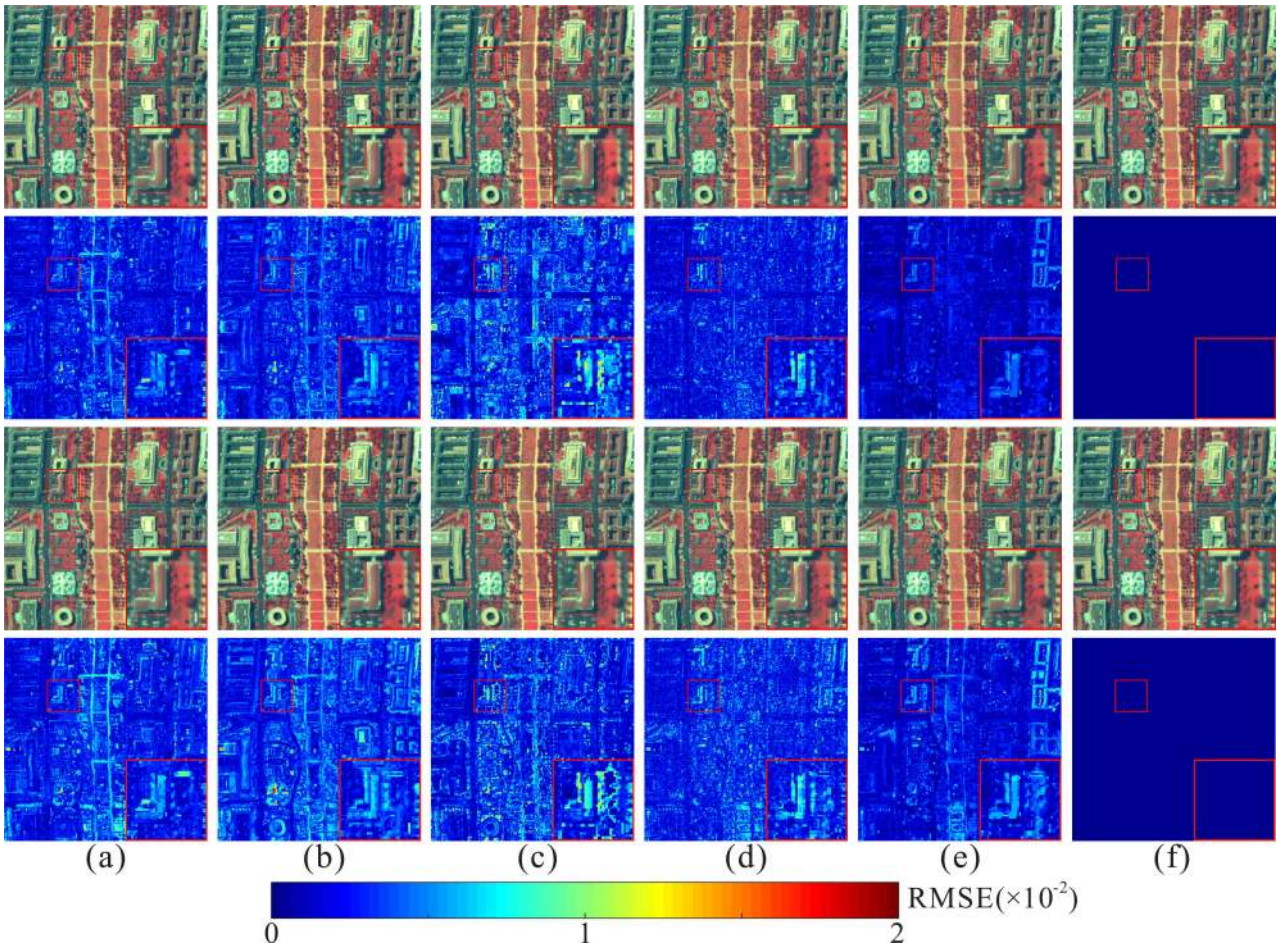

Fig. 5. Reconstructed images and the corresponding RMSE images of the Washington DC Mall dataset with uniform blur and scaling factor (Top two rows) $r = 8$ and (Bottom two rows) $r = 16$. (a) NSSR [26]. (b) CSTF [2]. (c) SSGLRTD [29]. (d) LTTR [30]. (e) Proposed method. (f) Reference image.

represents the SNR of the noisy LR-HSI $\mathcal{Y}$. The noisy HR-MSI $\mathcal{Z}$ is simulated by filtering the reference image using an IKONOS-like spectral reflectance response, and then adding the Gaussian white noise. We denote the SNR of the noisy HR-MSI $\mathcal{Z}$ by SNRm. The reconstructed images and the corresponding RMSE images of the Hyperspec Chikusei data from the LTTR and DHSIS methods and the proposed method are shown in Fig. 7. We only demonstrated the comparison with LTTR and DHSIS as they have better performances than other compared methods. Table IV presents the quality metric values of the noisy cases on the Hyperspec Chikusei dataset. We can clearly see that the proposed method still produces the best results compared to the compared methods in the noisy case.

TABLE IV
PERFORMANCE COMPARISON OF NOISY CASES ON THE HYPERSPEC CHIKUSEI DATASET WITH SCALING FACTOR 8

Method	NSSR	CSTF	SSGLRTD	LTTR	DHSIS	Proposed
SNRh = 35 dB, SNRm = 40 dB						
PSNR	41.548	43.471	41.899	45.376	44.310	**46.832**
SAM	1.956	2.030	1.929	1.406	1.844	**1.302**
DD	1.617	1.209	1.503	0.939	1.088	**0.818**
ERGAS	1.764	1.669	1.651	1.576	1.507	**1.409**
UIQI	0.964	0.964	0.967	0.970	0.970	**0.973**
SNRh = 30 dB, SNRm = 35 dB						
PSNR	40.631	42.280	40.392	43.846	43.142	**45.418**
SAM	2.225	2.243	2.370	1.669	2.033	**1.440**
DD	1.821	1.434	1.845	1.164	1.291	**0.983**
ERGAS	2.058	1.774	1.931	1.642	1.601	**1.475**
UIQI	0.951	0.957	0.950	0.964	0.964	**0.970**

E. Experimental Results on Real Dataset

In this section, we further conduct the real dataset experiment to verify the effectiveness of the proposed method and practical application in real-life scenarios. The real-life dataset consists of real LR-HSI data acquired from the Hyperion HSI sensor and real HR-MSI data acquired from Sentinel-2 satellite. The real data is obtained from Arkansas City, Cowley County, Kansas, USA. The used size of the real LR-HSI data is $100 \times 100 \times 89$. For the real HR-MSI data, only the R.G.B. and NIR bands of the original HR-MSI data are used and the used size of the real HR-MSI data is $300 \times 300 \times 4$ in this experiment. Thus, the spatial downsampling factor is 3. The spectral response matrix $\mathbf{P}$, spatial degradation method, and reference HR-HSI are unknown in this real dataset experiment. Therefore, we use the method in [22] to estimate the unknown spatial blur kernel and spectral response matrix. The CSTF is developed with the assumption that the spatial blur kernel can be decomposed into the width and height modes, while the blurring kernel estimated using the method developed in [22] is not separated. Thus, two

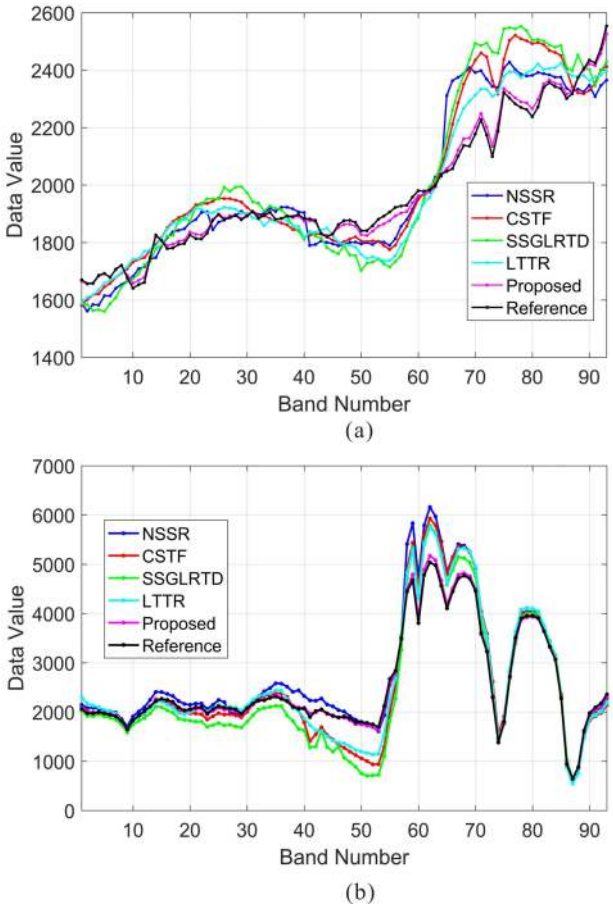

Fig. 6. Spectral curves of the reconstructed results with uniform blur and scaling factor $r = 8$. (a) Spectral curves of the [50,78]th pixels in the reconstructed University of Pavia HSI. (b) Spectral curves of the [225, 130]th pixels in the reconstructed Washington DC Mall HSI.

separate blur responses along the width and height modes are used to approximate the estimated spatial blur kernel in this experiment. In addition, considering that there is no reference HR-HSI in this experiment, the DHSIS is not implemented on the real dataset. Fig. 8 shows the reconstructed results of the test methods for the real dataset. The HR-HSI reconstructed by the NSSR shows obvious artifacts, while the HR-HSI reconstructed by the SSGLRTD exhibits spectral distortion. The HR-HSIs reconstructed by LTTR and CSTF exhibit obvious over-smoothing effects, indicating that the detailed spatial structures are not well reconstructed. However, the proposed method achieves better visual quality. Fig. 9 shows the spectral curves of the [130, 200]th pixel in the reconstructed results and the [50, 70]th pixel in the LR-HSI of the real dataset. We observe that the spectral signatures of the proposed method are more consistent with spectral signatures of the LR-HSI than the other compared methods. This experiment shows that the proposed method performs well in dealing with the superresolution problem of the real data.

V. Discussion

In this section, we provide more analysis of the proposed method through extensive experiments. For simplicity, only the

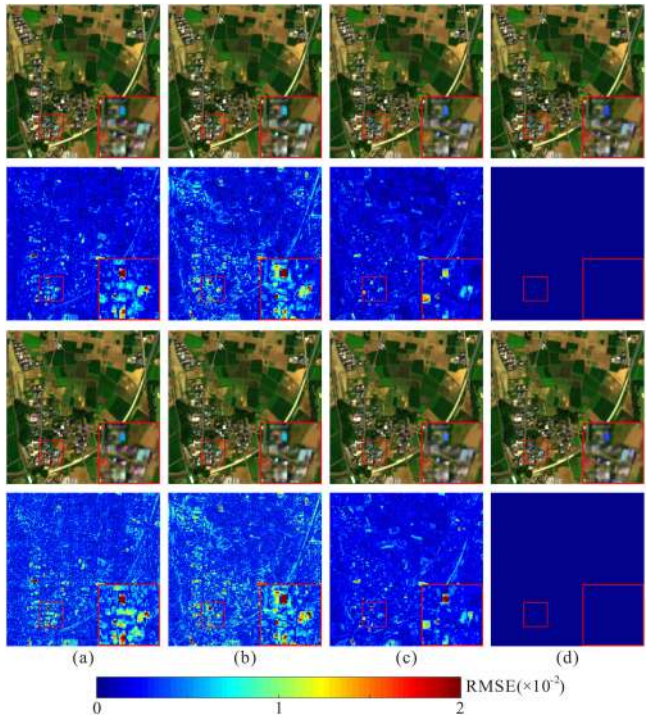

Fig. 7. Reconstructed images and the corresponding RMSE images of the Hyperspec Chikusei dataset with Gaussian white noises (Top two rows) SNRh = 35 dB and SNRm = 40 dB and (Bottom two rows) SNRh = 30 dB and SNRm = 35 dB. (a) LTTR [30]. (b) DHSIS [34]. (c) Proposed method. (d) Reference image.

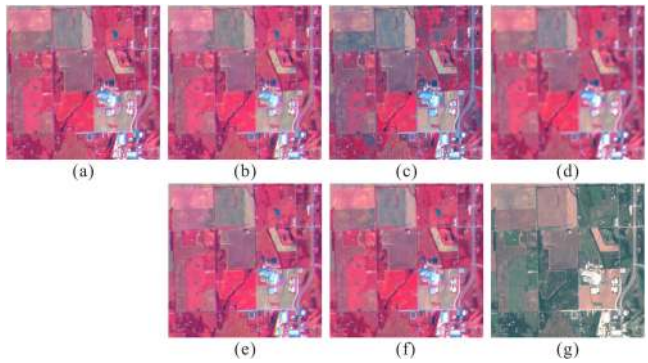

Fig. 8. Reconstructed images of the real data set. (a) NSSR [26]. (b) CSTF [2]. (c) SSGLRTD [29]. (d) LR-HSI. (e) LTTR [30]. (f) Proposed method. (g) HR-MSI.

University of Pavia dataset is selected as the experimental object for all following experiments. The LR-HSI $\mathcal{Y}$ is simulated by uniformly averaging over disjoint 8×8 blocks of the original University of Pavia data. The HR-MSI $\mathcal{Z}$ is generated by filtering the original University of Pavia data using the spectral reflectance response files of the R.G.B. and NIR bands of the IKONOS multispectral sensor.

A. *Comparison of Different* $\mathbf{E}_{(1)}$ *Terms*

In the proposed HL-GSNLTD method, we apply the nonconvex $\ell_{1/2}$ term to the gradients of the difference images of the upsampled LR-HSI and HR-HSI to enforce spectral preservation, which is only based on the empirical statistics of different

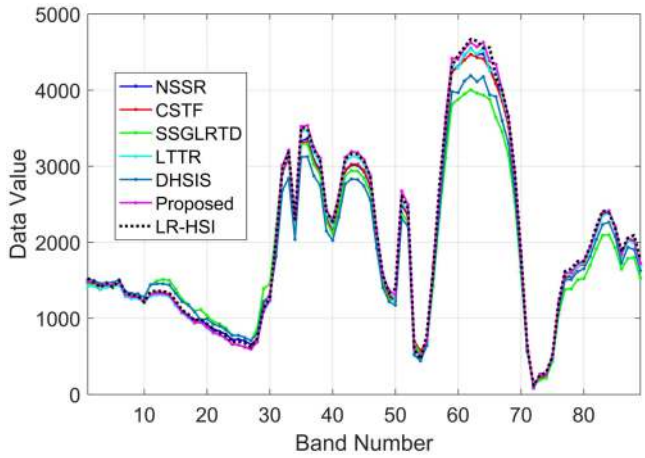

Fig. 9. Spectral curves of the [130, 200]th pixel in the reconstructed results and the [50, 70]th pixel in the LR-HSI of the real dataset.

TABLE V
NOTATION OF DIFFERENT $E_{(1)}$ TERMS

Notation	$E_{(1)}$ term
Proposed-ℓ_1	$\alpha\|\mathcal{X}-\hat{\mathcal{Y}}\|_1$
Proposed-ℓ_2	$\alpha\|\mathcal{X}-\hat{\mathcal{Y}}\|_2$
Proposed-ℓ_F	$\alpha\|\mathcal{Y}-\mathcal{X}\times_1\mathbf{P}_1\times_2\mathbf{P}_2\|_F^2$
Proposed-$\nabla\ell_1$	$\sum_{i=1}^{3}\alpha_i\|\nabla_i(\mathcal{X}-\hat{\mathcal{Y}})\|_1$
Proposed-$\nabla\ell_2$	$\sum_{i=1}^{3}\alpha_i\|\nabla_i(\mathcal{X}-\hat{\mathcal{Y}})\|_2$
Ours (*i.e.*, Proposed-$\nabla\ell_{1/2}$)	$\sum_{i=1}^{3}\alpha_i\|\nabla_i(\mathcal{X}-\hat{\mathcal{Y}})\|_{1/2}$

datasets. Here, by fixing the $\mathbf{E}_{(2)}$ and $\mathbf{E}_{(3)}$ terms, we investigate the performance with different $\mathbf{E}_{(1)}$ terms to the fusion model (7). The notations of $\mathbf{E}_{(1)}$ terms are summarized in Table V. It is worth noting that $\mathbf{P}_1$ and $\mathbf{P}_2$ in the Proposed-ℓ_F model are the blurring and downsampling matrices along the width and height modes, respectively. All these models are solved using ADMM-based algorithm. The 55th, 30th, and 8th band images of the reconstructed results are selected as a false color image shown in Fig. 10 for visual analysis. The first row in Fig. 10 displays the reference and the reconstructed images of all methods, and the second row displays the color-composite error images between the reference image and the reconstructed images. From Fig. 10, we obtain that the spatial details in all reconstructed images have been significantly enhanced. However, there are some spectral differences or blurring effects existing in the reconstructed images compared with the reference image. For example, the Proposed-ℓ_2 produces significant spectral distortion and fails to preserve the edges and details. The Proposed-ℓ_1, Proposed-ℓ_F, Proposed-$\nabla\ell_1$, and Proposed-$\nabla\ell_2$ perform well in spectral preservation, however, they have blurring and some loss of details. Compared with other methods, the proposed method (i.e., Proposed-$\nabla\ell_{1/2}$) provides a better visual quality. Table VI reports the quantitative evaluation results of the models with different $\mathbf{E}_{(1)}$ terms. The Proposed-ℓ_F is usually used as the spectral preservation term to deal with the HSI superresolution problem [2], [30], [65]. It takes second place in terms of quality metrics. The proposed method (i.e., Proposed-$\nabla\ell_{1/2}$) yields the best quantitative results on all quality measures, indicating the effectiveness of the proposed method.

TABLE VI
QUANTITATIVE EVALUATION OF THE MODELS WITH DIFFERENT $E_{(1)}$ TERMS ON THE UNIVERSITY OF PAVIA DATASET

Algorithm	PSNR	SAM	DD	ERGAS	UIQI
Best Values	∞	0	0	0	1
Proposed-ℓ_1	40.854	2.756	1.702	0.694	0.990
Proposed-ℓ_2	35.874	5.193	2.998	1.312	0.976
Proposed-ℓ_F	42.192	2.483	1.413	0.596	**0.993**
Proposed-$\nabla\ell_1$	42.128	2.495	1.462	0.605	0.992
Proposed-$\nabla\ell_2$	41.434	2.679	1.591	0.656	0.990
Proposed-$\nabla\ell_{1/2}$	**43.343**	**2.270**	**1.366**	**0.592**	**0.993**

B. Results With Different $\mathbf{E}_{(3)}$ Terms

The major contribution of the proposed HL-GSNLTD method is that both the nonlocal low-rank tensor decomposition and global SSTV regularization are introduced into the fusion model to simultaneously investigate the nonlocal low-rank property and global spatial-spectral smoothness of the HR-HSI. To investigate the influence of different $\mathbf{E}_{(3)}$ terms on the fusion results, we conduct additional comparisons among the models with different combinations of $\mathbf{E}_{(3)}$ terms, such as the basic fusion model (BFM=$\mathbf{E}_{(1)}$+$\mathbf{E}_{(2)}$), the basic fusion model with the nonlocal low rank prior (BFM+$\mathbf{E}_{(3)}^{\mathbf{NL}}$), the basic fusion model with the global SSTV prior (BFM+$\mathbf{E}_{(3)}^{\mathbf{TV}}$), and the basic fusion model with both the nonlocal low rank and global SSTV priors (i.e., the proposed model: BFM+$\mathbf{E}_{(3)}^{\mathbf{NL}}$+$\mathbf{E}_{(3)}^{\mathbf{TV}}$). For a fair comparison, all experimental environments and settings remain the same. From Fig. 11, we obtain that the BFM performs poorly at preserving the spatial and spectral information of the HR-HSI. After adding the nonlocal low-rank or global SSTV priors into the BFM, the quantitative performance increases significantly, indicating that $\mathbf{E}_{(3)}^{\mathbf{NL}}$ and $\mathbf{E}_{(3)}^{\mathbf{TV}}$ show important and positive influences on the fusion results. The proposed HL-GSNLTD model, which simultaneously considers the nonlocal low-rank and global SSTV priors, performs the best among the compared models. This indicates that the nonlocal low-rank and global SSTV priors can be organically combined to produce better results than considering only one of them.

C. Parameter Selection

In the proposed HL-GSNLTD method, several parameters with significant influences on the fusion results must be determined before performing Algorithm 2, including the number of cubes in the NSCT N_k for the cube clustering process, and some sensitive parameters for the optimization procedure, i.e., α_i, w_i ($i=1,2,3$), λ, μ, and p. Next, we take the University of Pavia dataset as an example and further investigate the influence of these sensitive parameters for our method.

The number of cubes in the NSCT N_k has a significant influence on the cube clustering process. Fig. 12(a) shows the PSNR of the reconstructed results of the University of Pavia dataset as a function of $N_k \in [2,16]$. As shown in Fig. 12(a), the PSNR values remain relatively stable for $N_k \in [2,8]$ and decrease sharply for $N_k > 8$. The PSNR attains its maximum value when $N_k = 6$. Thus, N_k is eventually set as 6. The parameters α_i ($i=1,2,3$) are the three regularization parameters for balancing the

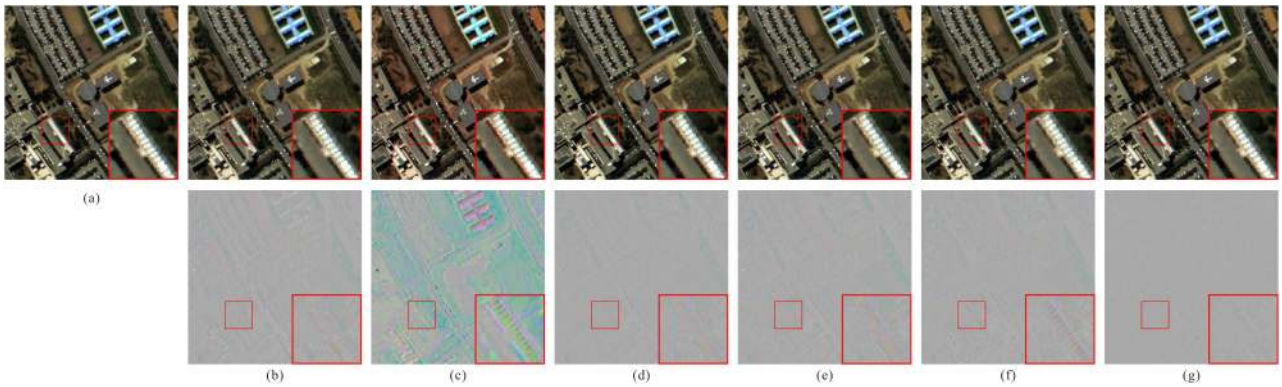

Fig. 10. Comparisons of the reconstructed results obtained using the models in Table V on the University of Pavia dataset. (a) Reference image. (b) Proposed-ℓ_1. (c) Proposed-ℓ_2. (d) Proposed-ℓ_F. (e) Proposed-$\nabla\ell_1$. (f) Proposed-$\nabla\ell_2$. (g) Proposed-$\nabla\ell_{1/2}$.

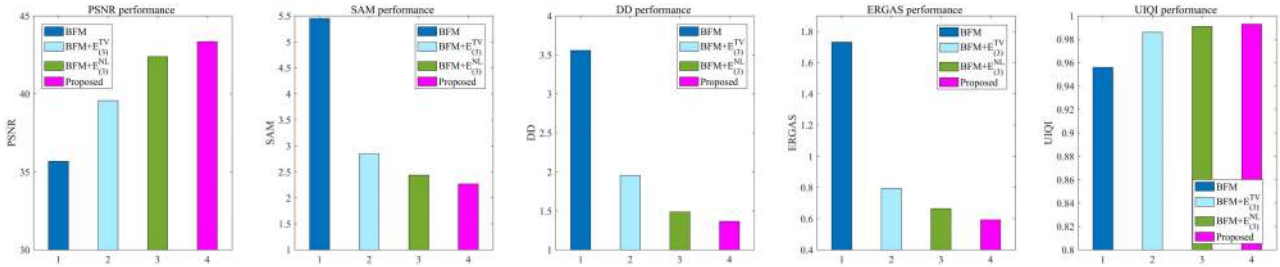

Fig. 11. Quantitative evaluation of the models with different combinations of $\mathbf{E}_{(3)}$ terms, i.e., BFM=$\mathbf{E}_{(1)}+\mathbf{E}_{(2)}$, BFM+$\mathbf{E}_{(3)}^{\mathbf{TV}}$, BFM+$\mathbf{E}_{(3)}^{\mathbf{NL}}$, and the proposed model (i.e., BFM+$\mathbf{E}_{(3)}^{\mathbf{NL}}+\mathbf{E}_{(3)}^{\mathbf{TV}}$), for the University of Pavia dataset. The proposed model exhibits the best performance compared to other models.

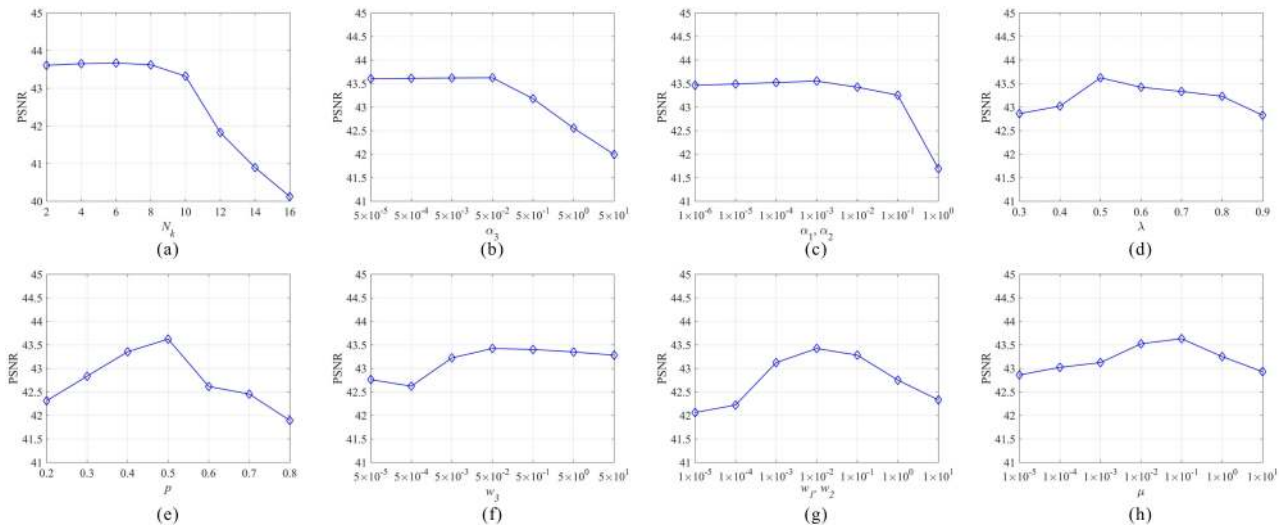

Fig. 12. PSNR curves as a function of the parameters for the proposed method. (a) Number of cubes N_k. (b) Parameter α_3. (c) Parameters α_i $(i=1,2)$. (d) Parameter λ. (e) Parameter p. (f) Parameter w_3. (g) Parameters w_i $(i=1,2)$. (h) Parameter μ.

contribution of the nonconvex ℓ_p term in the different directions. From Fig. 12(b) and (c), we can clearly see that the proposed method achieves high PSNR measures when $\alpha_3 \in [5\times 10^{-5}, 5\times 10^{-2}]$ and $\alpha_1=\alpha_2 \in [1\times 10^{-6}, 1\times 10^{-2}]$. Thus, α_1 and α_2 can be set to 1×10^{-3}, and α_3 to 5×10^{-2}. Fig. 12(d), (e), and (h) shows the PSNR curves of the reconstructed results for the parameters λ, p and μ, respectively. We obtain that when $\lambda=0.5$, $p=0.5$, and $\mu=1\times 10^{-1}$, the PSNR measures attain their maximum values, respectively. Thus, we set $\lambda=p=0.5$ and $\mu=1\times 10^{-1}$. The parameters w_i $(i=1,2,3)$ are the three regularization parameters for balancing the contribution of the spatial-spectral TV term in the different directions. Fig. 12(f) and (g) shows the PSNR of reconstructed results as a function of w_3 and $w_1=w_2$, respectively. It shows that the proposed method achieves the best PSNR performance when $w_3=5\times 10^{-2}$ and $w_1=w_2=1\times 10^{-2}$, respectively. Thus, in our experiments, we set $w_3=5\times 10^{-2}$ and $w_1=w_2=1\times 10^{-2}$. Besides, we test the proposed method when the penalty parameters δ, η_i

TABLE VII
TIME COMPARISON OF DIFFERENT METHODS ON DIFFERENT DATASETS (IN SECONDS)

Data Set	NSSR [26]	CSTF [2]	SSGLRTD [29]	LTTR [30]	DHSIS [34]	Proposed
University of Pavia	50.3	40.7	763.4	455.2	–	301.7
Washington DC Mall	49.4	41.3	761.8	456.3	–	301.7
Hyperspec Chikusei	51.2	40.5	762.5	455.5	18.3	301.5

and γ_i $(i = 1, 2, 3)$ varies from 1×10^{-5} to 1×10^{1}. Then, it can be set as $\delta = \eta_3 = \gamma_3 = 1 \times 10^{-3}$, and $\eta_i = \gamma_i = 1 \times 10^{-2}$ $(i = 1, 2)$ to achieve the best results on the University of Pavia dataset.

D. Computational Complexity

In this article, all experiments were implemented using MATLAB R2016b on a computer with an Intel(R) Core(TM) i7-8750H CPU @ 2.21 GHz and 16 GB of RAM. For each iteration of the proposed HL-GSNLTD method, the computational burden mainly consists of three parts, i.e., the nonlocal low-rank tensor decomposition, global SSTV regularization related subproblem, and nonconvex sparse ℓ_p related subproblem. The nonlocal low-rank tensor decomposition can be accelerated by the singular value thresholding operator after applying the linearizing operation on the quadratic term. The global SSTV regularization related subproblem can be accurately and quickly solved by the FFT algorithm. The nonconvex sparse ℓ_p subproblem is readily solved using the fast GST algorithm. Table VII presents the computational times of different methods on the three datasets. We observe that the proposed method consumes more time than NSSR, CSTF, and DHSIS; however, it is less time consuming than the popular local and nonlocal tensor-based methods, such as SSGLRTD and LTTR. Although DHSIS is very fast at predicting the HR-HSI, it requires long computational time on the neural network training process.

VI. CONCLUSION

In this article, a novel global gradient sparse and nonlocal low-rank tensor decomposition model with a hyper-Laplacian prior is proposed for fusing the LR-HSI and HR-MSI of the same scene. The proposed model mainly consists of three terms, i.e., the spectral and spatial preservation terms and the prior knowledge term. The spectral preservation term is described as an ℓ_p $(0 < p < 1)$ energy term, which was inspired by the investigated hyper-Laplacian distribution of the gradients of the difference images between the upsampled LR-HSI and latent HR-HSI. The spatial preservation term is described as a tensor-based fidelity term to recover the spatial details. The last term models both the nonlocal low-rank structure and the global spatial piecewise smooth and spectral consistent structures of the HR-HSI. Specifically, the nonlocal low rank Tucker decomposition model is used to learn the spatial and spectral similarities among the nonlocal similar cubes in the HR-HSI. The global SSTV regularization is adopted to reconstruct these nonlocal similar patches, and to further capture the global spatial piecewise smoothness and spectral consistency of the HR-HSI. The proposed optimization problem is efficiently solved using ADMM. The proposed method was visually and quantitatively compared with other existing state-of-the-art fusion methods on three synthetic data sets and one real-life dataset, verifying the superiority of our method. In addition, more analysis about the influences of parameters and energy terms are also presented, and the results demonstrate the effectiveness of our method.

In the future works, we plan to explore the parallel acceleration scheme of the proposed method. In addition, tensor-based techniques are much more advantageous in higher dimensional image processing compared to matrix-based techniques [2], [29]–[31]. Thus, we will investigate the feasibility of other tensor techniques, such as tensor ring decomposition [66], to improve the reconstruction results.

REFERENCES

[1] 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.

[2] 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.

[3] R. Dian, L. Fang, and S. Li, "Hyperspectral image super-resolution via non-local sparse tensor factorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2017, pp. 5344–5353.

[4] 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.

[5] G. A. Shaw and H. K. Burke, "Spectral imaging for remote sensing," *Lincoln Lab. J.*, vol. 14, no. 1, pp. 3–28, Jan. 2003.

[6] C. Thomas, T. Ranchin, L. Wald, and J. Chanussot, "Synthesis of multispectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1301–1312, May 2008.

[7] G. Vivone *et al.*, "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, May 2015.

[8] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "Twenty-Five Years of Pansharpening: A Critical Review and New Developments," *in Signal and Image Processing for Remote Sensing*. Boca Raton, FL, USA: CRC Press, 2012, pp. 552–599.

[9] 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.

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

[11] Q. Yuan, Y. Wei, X. Meng, H. Shen, and L. Zhang, "A multiscale and multidepth convolutional neural network for remote sensing imagery pansharpening," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 978–989, Mar. 2018.

[12] L.-J. Deng, M. Feng, and X.-C. Tai, "The fusion of panchromatic and multispectral remote sensing images via tensor-based sparse modeling and hyper-laplacian prior," *Inf. Fusion*, vol. 52, pp. 76–89, Dec. 2019.

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

[14] X. Meng, H. Shen, H. Li, L. Zhang, and R. Fu, "Review of the pansharpening methods for remote sensing images based on the idea of meta-analysis: Practical discussion and challenges," *Inf. Fusion*, vol. 46, pp. 102–113, Mar. 2019.

[15] H. Zhang and B. Huang, "A new look at image fusion methods from a Bayesian perspective," *Remote Sens.*, vol. 7, no. 6, pp. 6828–6861, May 2015.

[16] M. Moeller, T. Wittman, and A. L. Bertozzi, "A variational approach to hyperspectral image fusion," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XV*, vol. 7334, International Society for Optics and Photonics, Apr. 2009, p. 73341E.

[17] L. Alparone, B. Aiazzi, S. Baronti, and A. Garzelli, *Remote Sensing Image Fusion*. Boca Raton, FL, USA: CRC Press, 2015.

[18] G. Vivone, R. Restaino, G. Licciardi, M. Dalla Mura, and J. Chanussot, "Multiresolution analysis and component substitution techniques for hyperspectral pansharpening," in *Proc. IEEE Geosci. Remote Sens. Symp.*, May 2014, pp. 2649–2652.

[19] M. Cetin and N. Musaoglu, "Merging hyperspectral and panchromatic image data: Qualitative and quantitative analysis," *Int. J. Remote Sens.*, vol. 30, no. 7, pp. 1779–1804, Apr. 2009.

[20] 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.

[21] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Bayesian fusion of multi-band images," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1117–1127, Sep. 2015.

[22] 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.

[23] 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., Jun. 2011, pp. 2329–2336.

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

[25] Q. Wei, J. Bioucas-Dias, N. Dobigeon, J.-Y. Tourneret, M. Chen, and S. Godsill, "Multiband image fusion based on spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7236–7249, Dec. 2016.

[26] 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.

[27] O. Berné, A. Helens, P. Pilleri, and C. Joblin, "Non-negative matrix factorization pansharpening of hyperspectral data: An application to mid-infrared astronomy," in *Proc. IEEE 2nd Workshop Hyperspectral Image Signal Process., Evolution Remote Sens.*, Jun. 2010, pp. 1–4.

[28] E. Wycoff, T.-H. Chan, K. Jia, W.-K. Ma, and Y. Ma, "A non-negative sparse promoting algorithm for high resolution hyperspectral imaging," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Oct. 2013, pp. 1409–1413.

[29] 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.

[30] 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.

[31] 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.

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

[33] Y. Li, J. Hu, X. Zhao, W. Xie, and J. Li, "Hyperspectral image super-resolution using deep convolutional neural network," *Neurocomputing*, vol. 266, pp. 29–41, Nov. 2017.

[34] 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.

[35] Y. Qu, H. Qi, and C. Kwan, "Unsupervised sparse dirichlet-net for hyperspectral image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2511–2520.

[36] Y. Yuan, X. Zheng, and X. Lu, "Hyperspectral image superresolution by transfer learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 1963–1974, May 2017.

[37] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. Adv. Neural Inf. Process. Syst.*, Feb. 2001, pp. 556–562.

[38] W. He, H. Zhang, H. Shen, and L. Zhang, "Hyperspectral image denoising using local low-rank matrix recovery and global spatial-spectral total variation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 713–729, Mar. 2018.

[39] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D, Nonlinear Phenomena*, vol. 60, no. 1–4, pp. 259–268, Nov. 1992.

[40] Y. Wang, J. Peng, Q. Zhao, Y. Leung, X.-L. Zhao, and D. Meng, "Hyperspectral image restoration via total variation regularized low-rank tensor decomposition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1227–1243, Apr. 2018.

[41] H. Fan, C. Li, Y. Guo, G. Kuang, and J. Ma, "Spatial-spectral total variation regularized low-rank tensor decomposition for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 6196–6213, Oct. 2018.

[42] H. Zhang, L. Liu, W. He, and L. Zhang, "Hyperspectral image denoising with total variation regularization and nonlocal low-rank tensor decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3071–3084, May 2020.

[43] S. Boyd *et al.*, "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.

[44] W. Zuo, D. Meng, L. Zhang, X. Feng, and D. Zhang, "A generalized iterated shrinkage algorithm for non-convex sparse coding," in *Proc. IEEE Int. Conf. Comput. Vis.*, Mar. 2013, pp. 217–224.

[45] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, Aug. 2009.

[46] 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, Jul. 2019.

[47] Q. Wei, N. Dobigeon, and J.-Y. 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.

[48] Q. Wei, N. Dobigeon, J.-Y. Tourneret, J. Bioucas-Dias, and S. Godsill, "R-fuse: Robust fast fusion of multiband images based on solving a Sylvester equation," *IEEE Signal Process. Lett.*, vol. 23, no. 11, pp. 1632–1636, Nov. 2016.

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

[50] C. Zhao, X. Gao, W. J. Emery, Y. Wang, and J. Li, "An integrated spatio-spectral-temporal sparse representation method for fusing remote-sensing images with different resolutions," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3358–3370, Jun. 2018.

[51] W. Cao *et al.*, "Total variation regularized tensor RPCA for background subtraction from compressive measurements," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4075–4090, Sep. 2016.

[52] Q. Xie *et al.*, "Multispectral images denoising by intrinsic tensor sparsity regularization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Dec. 2016, pp. 1692–1700.

[53] I. Ram, M. Elad, and I. Cohen, "Image processing using smooth ordering of its patches," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2764–2774, Jul. 2013.

[54] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 208–220, Jan. 2013.

[55] J. Yang and X. Yuan, "Linearized augmented lagrangian and alternating direction methods for nuclear norm minimization," *Math. Comput.*, vol. 82, no. 281, pp. 301–329, Mar. 2013.

[56] J.-F. Cai, E. J. Candés, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, Mar. 2010.

[57] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2008, pp. 3869–3872.

[58] Y. She *et al.*, "Thresholding-based iterative selection procedures for model selection and shrinkage," *Electron. J. Statist.*, vol. 3, pp. 384–415, Dec. 2009.

[59] Y. She, "An iterative algorithm for fitting nonconvex penalized generalized linear models with grouped predictors," *Comput. Statist. Data Anal.*, vol. 56, no. 10, pp. 2976–2990, Oct. 2012.

[60] D. Krishnan and R. Fergus, "Fast image deconvolution using hyper-laplacian priors," in *Proc. Adv. Neural Inf. Process. Syst.*, Jan. 2009, pp. 1033–1041.

[61] J. M. Nascimento and J. M. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.
[62] R. H. Yuhas, A. F. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm," in *Proc. Summaries 4th JPL Airborne Earth Sci. Workshop*, Dec. 1992, pp. 147–149.
[63] L. Wald, "Quality of high resolution synthesised images: Is there a simple criterion?" in *Proc. 3rd Conf. "Fusion of Earth data: Merging Point Measurements, Raster Maps and Remotely Sensed Images*," Jan. 2000, pp. 99–103.
[64] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.
[65] W. Wan, W. Guo, H. Huang, and J. Liu, "Nonnegative and nonlocal sparse tensor factorization-based hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8384–8394, Dec. 2020.
[66] Y. Chen, W. He, N. Yokoya, T.-Z. Huang, and X.-L. Zhao, "Nonlocal tensor-ring decomposition for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 2, pp. 1348–1362, Feb. 2020.

Yidong Peng was born in Chongqing, P.R. China, in 1988. He received the M.S. degree in Computer Science and Technology from the Institute of Computer science and technology, Chongqing University of Posts and Telecommunications, Chongqing, P.R. China, in July 2017. Currently, he is working toward the Ph.D. degree in computer science and technology at the Chongqing Key Laboratory of Image Cognition, Chongqing University of Posts and Telecommunications.

His research interests include urban thermal infrared remote sensing, remote sensing image processing, and ecological environment monitoring and evaluating.

Weisheng Li received his B.S. degree in Industrial Automation from the School of Electronics and Mechanical Engineering at Xidian University, Xian, China, in July 1997. He received his M.S. degree in Mechanical manufacturing and automation from the School of Electronics and Mechanical Engineering at Xidian University in Mar 2000. He received his Ph.D. degree in Computer Science and Technology from the School of Computer Science and Technology at Xidian University in July 2004. Currently, he is a professor at Chongqing University of Posts and Telecommunications. His current research focuses on intelligent information processing and pattern recognition.

Xiaobo Luo received the B.S. degree in GIS from the Institute of Geophysics and Geomatics, China University of Geosciences, Wuhan, China, in 1999, the M.S. degree in GIS from the School of Information Engineering, Chinese Academy of Science, Wuhan, China, in 2004, and the Ph.D. degree in cartography and GIS from the Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, China, in 2010.

He is currently a Professor with the Chongqing University of Posts and Telecommunications, Chongqing, China. His current research interests include urban thermal infrared remote sensing, remote sensing image processing, and ecological environment monitoring and evaluation.

Jiao Du was born in Chongqing, P.R. China, in 1988. She received M.S. degree and Ph.D. degree in Computer Science and Technology from the Institute of Computer science and technology, Chongqing University of Posts and Telecommunications, Chongqing, P.R. China in July 2013 and in July 2017, respectively.

Currently, she is a Lecture of School of Computer Science and Cyber Engineering, Guangzhou University, Guangzhou, China. Her research interests include pattern recognition and image processing.