Hyperspectral Image Restoration Using Weighted Group Sparsity-Regularized Low-Rank Tensor Decomposition

Yong Chen, Wei He, *Member, IEEE*, Naoto Yokoya, *Member, IEEE*, and Ting-Zhu Huang

Abstract—Mixed noise (such as Gaussian, impulse, stripe, and deadline noises) contamination is a common phenomenon in hyperspectral imagery (HSI), greatly degrading visual quality and affecting subsequent processing accuracy. By encoding sparse prior to the spatial or spectral difference images, total variation (TV) regularization is an efficient tool for removing the noises. However, the previous TV term cannot maintain the shared group sparsity pattern of the spatial difference images of different spectral bands. To address this issue, this article proposes a group sparsity regularization of the spatial difference images for HSI restoration. Instead of using ℓ_1- or ℓ_2-norm (sparsity) on the difference image itself, we introduce a weighted $\ell_{2,1}$-norm to constrain the spatial difference image cube, efficiently exploring the shared group sparse pattern. Moreover, we employ the well-known low-rank Tucker decomposition to capture the global spatial–spectral correlation from three HSI dimensions. To summarize, a weighted group sparsity-regularized low-rank tensor decomposition (LRTDGS) method is presented for HSI restoration. An efficient augmented Lagrange multiplier algorithm is employed to solve the LRTDGS model. The superiority of this method for HSI restoration is demonstrated by a series of experimental results from both simulated and real data, as compared with the other state-of-the-art TV-regularized low-rank matrix/tensor decomposition methods.

Index Terms—Augmented Lagrange multiplier (ALM) algorithm, group sparsity, hyperspectral image restoration, low-rank tensor decomposition.

I. INTRODUCTION

HYPERSPECTRAL imagery (HSI) is collected by imaging spectroscopy over hundreds of bands and has rich spectral information from all bands compared with that of color or multispectral images. Due to its high-spectral resolution, HSI has various applications in remote sensing, such as classification [1], unmixing [2], and target detection [3]. HSI is typically degraded by a variety of mixed noises in a real scene because of facility restrictions and weather conditions during the collection process [4]. The noise in HSI seriously degrades the visual quality and affects further processing accuracy. Therefore, restoring the noisy HSI is an important issue to address prior to subsequent applications.

During the past few decades, many different HSI restoration approaches have been proposed to improve its quality. A straightforward technique is to use a conventional 2-D grayscale image and 1-D signal noise removal approaches for HSI band-by-band [5] or pixel-by-pixel [6]. However, these methods cannot achieve satisfactory restoration results because they fail to make full use of the strong correlation among all bands and adjacent pixels. To address this issue, more efficient approaches have been proposed by combining HSI correlation along the spatial–spectral directions. For example, utilizing the difference in signal regularity across the spatial–spectral modes, a spatial–spectral derivative-domain wavelet shrinkage method has been proposed for HSI restoration [7]. In addition, the sparse representation methods [8] have also been proposed to efficiently encode the spatial–spectral information for HSI restoration.

In real scenes, HSI is often degraded by a variety of mixed noise during the collection process, such as Gaussian, impulse, stripe, deadlines, and many unknown noises [4]. Recently, the robust principal component analysis framework (RPCA) [9] has been widely utilized for mixed noise removal. Typically, HSI is assumed to have the low-rank property along the spectral dimension [10], [11]. Meanwhile, deadlines, stripes, and impulse noises have a sparse property [4]. Given these assumptions, classical low-rank matrix recovery (LRMR) was first presented by Zhang *et al.* [4] for HSI mixed noise removal. To improve the HSI restoration performance of LRMR, many nonconvex low-rank matrix approximation functions, such as γ-norm [12] and weighted Schatten p-norm [13], [14], have been proposed to formulate the low-rank approximation instead of the original

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

2   IEEE TRANSACTIONS ON CYBERNETICS
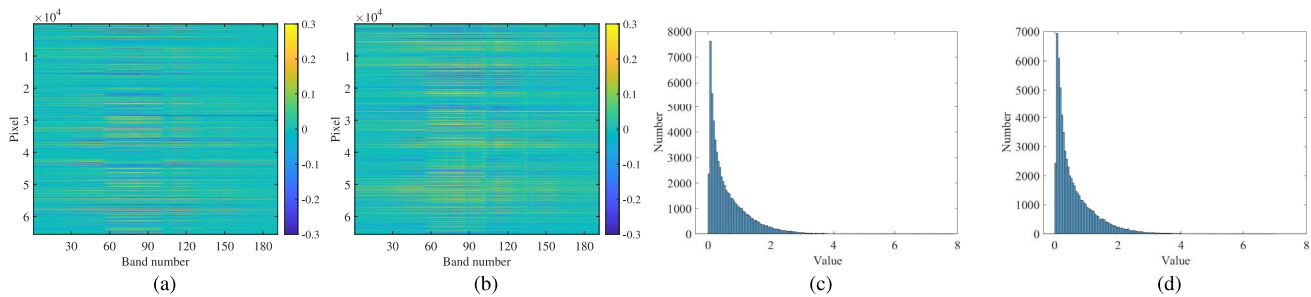

Fig. 1. (a) Difference image of spatial horizontal dimension. (b) Difference image of spatial vertical dimension (zoom in for better visibility). (c) Histogram distribution of (a). (d) Histogram distribution of (b).

rank minimization. Moreover, low-rank matrix factorization was also applied for HSI mixed noise removal [15]–[17] to avoid computation of singular value decomposition and improve the efficiency. Low-rank regularization on the spectral difference image [18] and superpixel-based subspace low-rank representation [19] were proposed for HSI restoration. To take full advantage of the deep convolution neural networks in image processing, deep learning (DL)-based methods were also widely applied to HSI restoration [20]–[22]. The low-rank matrix-based methods need to reshape the original 3-D HSI into a 2-D matrix and ignore the HSI spatial correlation. Therefore, the low-rank tensor-based methods have been introduced to describe the HSI low-rank property, including Tucker decomposition [23], [24]; parallel factor analysis (PARAFAC) decomposition [25], [26]; and tensor singular value decomposition (t-SVD) [27], [28]. However, as presented in [29], the low-rank regularization is not sufficient to describe the prior spatial information.

Total variation (TV) regularization is a well-known tool to preserve the local spatial piecewise smoothness in image processing [30]–[33]. Therefore, many researchers have introduced various types of TV regularization based on the low-rank matrix/tensor decomposition framework to simultaneously explore the spatial and spectral priors for HSI. First, He *et al.* [29] introduced band-by-band TV regularization into the low-rank matrix factorization framework to boost the restoration results. Subsequently, by expanding the band-by-band TV regularization, spatial–spectral TV regularization has been widely adopted to capture spatial and spectral smoothness [34], [35]. For example, spatial–spectral TV regularization with a weighted nuclear norm [36], local low-rank regularization [37], spectral difference low-rank regularization [38], t-SVD [39], and low-rank Tucker decomposition [40]. Moreover, vectorial TV [41] and 3-D spectral–spatial cross TV [42] have also been introduced to low-rank models for HSI mixed noise removal. In summary, the combination of a low-rank tensor framework and spatial–spectral TV regularization can achieve state-of-the-art results.

Although existing TV regularization with low-rank matrix/tensor modeling has achieved satisfactory HSI restoration results, it still poses one problem. Fig. 1 presents the difference image and histogram distribution along the spectral dimension for Washington DC Mall data in both spatial dimensions. Fig. 1(a) and (b) shows the difference images
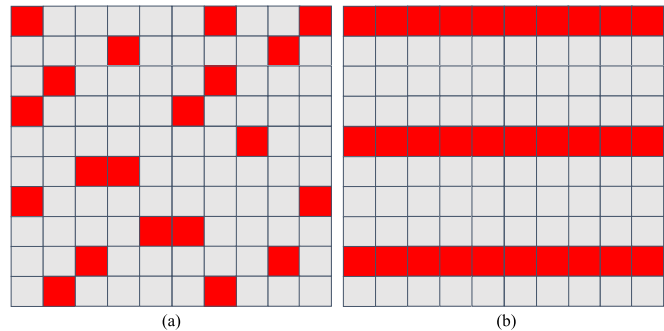

Fig. 2. Illustration of (a) sparsity and (b) group sparsity. The gray squares represent the zero elements and the red squares represent the nonzero elements. In (a), the elements show a random distribution. In (b), the elements show a group distribution.

of the entire scene. It was found that the elements of each row approximately tend to be all zero or all nonzero and the number of zero rows is much larger than that of the nonzero rows, indicating a shared group sparse pattern for these elements. Fig. 1(c) and (d) shows the histogram distribution of the ℓ_2-norm value of each row (contains 191 numbers) for Fig. 1(a) and (b), respectively. It is clear that most ℓ_2-norm values are zero or near zero. Previous TV methods simply adopted ℓ_1-norm to regularize the sparse prior of the spatial difference images, but these methods cannot exploit the shared row sparse structure of the difference images. The differences between the sparsity and group sparsity are shown in Fig. 2. From the figure, we can observe that Fig. 2(b) has an obvious row sparse structure compared with the sparse structure as shown in Fig. 2(a). Thus, this leads one considers the group sparsity regularization instead of the sparsity regularization to promote the inner group relationships among the spatial difference images of the different bands.

In this article, we propose a group sparsity-regularized TV combined with low-rank Tucker decomposition for HSI restoration. We used group sparsity regularization, denoted by $\ell_{2,1}$-norm, to explore the shared sparse pattern of a difference image in both spatial dimensions, which is totally different from the previous TV regularization [43]. Moreover, a weighted strategy was also introduced to promote the group sparsity along the spectral dimension. The proposed weighted group sparsity term was embedded into the low-rank Tucker

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

CHEN *et al.*: HYPERSPECTRAL IMAGE RESTORATION USING WEIGHTED GROUP SPARSITY-REGULARIZED LOW-RANK TENSOR DECOMPOSITION 3

decomposition framework. The contributions of this article are summarized as follows.

1) We proposed a weighted group sparsity regularization, which is represented by $\ell_{2,1}$-norm, to better capture the shared sparse pattern of the difference images for different bands in both spatial dimensions, which is expected to improve the restoration results compared to those of the previous TV methods.

2) We incorporated the weighted group sparsity term into the low-rank Tucker decomposition model for HSI mixed noise removal. Low-rank Tucker decomposition was utilized to separate the desired HSI and preserve the spatial and spectral correlation across all HSI bands.

3) We designed an augmented Lagrange multiplier (ALM) algorithm to solve our weighted group sparsity-regularized low-rank Tucker decomposition model; a series of experimental results demonstrate the superiority of the proposed method compared with the other state-of-the-art methods based on the combination of TV regularization and low-rank matrix/tensor decomposition methods.

The remainder of this article is organized as follows. Some tensor notations and preliminaries of tensors and related HSI restoration methods are presented in Section II. Section III introduces the proposed weighted group sparsity-regularized low-rank tensor decomposition model and its optimization procedure. A series of experiments with both simulated and real data and discussions are reported in Section IV. Finally, Section V concludes this article.

II. TENSOR NOTATIONS AND RELATED WORK

A. Notations and Preliminaries

In this section, we introduce some notations and preliminaries for tensor. Tensor is a multidimensional data array. The capitalized calligraphic letter was used to denote tensor data, for example, $\mathcal{Y}$. We used capitalized boldface letter to denote matrices (2-D data), for example, $\mathbf{Y}$. A vector (1-D data) was represented as a lowercase boldface letter, for example, $\mathbf{y}$, and a scalar was represented as a lowercase letter, for example, y. An N-dimensional tensor in a real number was denoted as $\mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, and $y_{i_1,i_2,\ldots,i_N}$ is its $(i_1, i_2, \ldots, i_N)$ element.

The fiber of a tensor was formed by fixing all but one index. A 3-D tensor contains row, column, and tube fibers, respectively, defined by $\mathbf{y}_{i:k}$, $\mathbf{y}_{:jk}$, and $\mathbf{y}_{ij:}$. A 2-D section of tensor was represented as a slice, denoted by fixing every index but two. The lateral, horizontal, and frontal slides of a 3-D tensor $\mathcal{Y}$ were represented by $\mathbf{Y}_{:j:}$, $\mathbf{Y}_{i::}$, and $\mathbf{Y}_{::k}$, respectively. $\mathbf{Y}_{(n)}$ was the mode-n unfolding of tensor $\mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, formed by arranging the mode-n fibers as columns [44].

The calculation of the inner product for two tensors $\mathcal{Y}_1$ and $\mathcal{Y}_2$ was $\langle \mathcal{Y}_1, \mathcal{Y}_2 \rangle = \sum_{i_1,i_2,\ldots,i_N} y_{1_{i_1,i_2,\ldots,i_N}} \cdot y_{2_{i_1,i_2,\ldots,i_N}}$. The Frobenius norm of tensor $\mathcal{Y}$ was computed by $\|\mathcal{Y}\|_F = \sqrt{\langle \mathcal{Y}, \mathcal{Y} \rangle}$. Moreover, the ℓ_1-norm of tensor $\mathcal{Y}$ was calculated as $\|\mathcal{Y}\|_1 = \sum_{i_1,i_2,\ldots,i_N} |y_{i_1,i_2,\ldots,i_N}|$. The mode-$n$ multiplication of the tensor and matrix was defined as $(\mathcal{Y} \times_n \mathbf{U})_{i_1,\ldots,i_{n-1}i_{n+1},\ldots,i_N} = \sum_{i_n} y_{i_1,i_2,\ldots,i_N} \cdot u_{j,i_n}$. The multilinear rank of a tensor was a vector $(r_1, r_2, \ldots, r_N)$, where $r_n = \text{rank}(\mathbf{Y}_{(n)})$, $n = 1, 2, \ldots, N$. Please refer to [44] for a more detailed introduction of tensors.

B. Problem Formulation

A contaminated HSI is typically degraded by additive mixed noise, mostly containing Gaussian, impulse, stripe, and deadline noises [4]. Therefore, the HSI mixed noise degradation model was formulated as follows:

$$\mathcal{Y} = \mathcal{X} + \mathcal{S} + \mathcal{N}. \tag{1}$$

In model (1), $\mathcal{Y}$ is the noisy HSI; $\mathcal{X}$ is the latent clean HSI; $\mathcal{S}$ denotes the sparse noise, mainly containing a mixture of impulse, stripe, and deadline noises; and $\mathcal{N}$ is the Gaussian noise. The four components $\mathcal{Y}, \mathcal{X}, \mathcal{S}$, and $\mathcal{N}$ are 3-D tensors with the same size of $m \times n \times b$, where $m \times n$ denotes the spatial size of each band and b denotes the number of spectral bands. HSI restoration focuses on how to restore a latent image $\mathcal{X}$ from an observed noisy image $\mathcal{Y}$.

C. Low-Rank Restoration of HSI

Recovering $\mathcal{X}$ from noisy $\mathcal{Y}$ is a difficult ill-posed problem. To address this problem, the prior regularization of different unknown variables is necessary. Therefore, optimization of the mixed noise removal problem can be formulated as follows:

$$\min_{\mathcal{X},\mathcal{S}} J_1(\mathcal{X}) + \lambda J_2(\mathcal{S}) \quad \text{s.t.} \quad \|\mathcal{Y} - \mathcal{X} - \mathcal{S}\|_F^2 \leq \varepsilon \tag{2}$$

where $J_1(\mathcal{X})$ and $J_2(\mathcal{S})$ are the regularization terms describing the prior knowledge of the clean HSI $\mathcal{X}$ and sparse noise $\mathcal{S}$, respectively. λ is the positive regularization parameter and ε is the Gaussian noise density variance. Typically, $J_2(\mathcal{S})$ is assumed to be sparse regularization for the sparse noise [4], that is, $J_2(\mathcal{S}) = \|\mathcal{S}\|_1$. Next, we show how we investigated a formula for the regularization term $J_1(\mathcal{X})$.

According to the linear mixture model [10], HSI possesses a strong correlation in the spectral dimension, indicating that the matrix $\mathbf{X}_{(3)}$ has a low-rank prior. Therefore, different types of low-rank matrix approximation regularization are utilized to explore the prior information of $J_1(\mathcal{X})$.

The low-rank matrix restoration methods make the vectorization of all HSI bands. However, this step loses the correlation of the spatial structures. Because HSI is a three-order tensor, it is more reasonable to adopt tensor tools to model it for preserving details and spatial structures. Wang *et al.* [40] have claimed that HSI also has a specific correlation in both spatial dimensions by observing the obvious decaying trends of the singular value changes of $\mathbf{X}_{(1)}$ and $\mathbf{X}_{(2)}$. To simultaneously aggregate the spatial–spectral correlation of HSI in three dimensions, efficient Tucker decomposition can be employed to constrain the clean HSI $\mathcal{X}$, that is, $J_1(\mathcal{X}) = \|\mathcal{X} - \mathcal{G} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3\|_F^2$. By choosing a proper regularization parameter, the regularization model (2) is equivalent to the following optimization problem under the low-rank Tucker decomposition constraint:

$$\min_{\mathcal{X},\mathcal{S}} \lambda \|\mathcal{S}\|_1 \quad \text{s.t.} \quad \mathcal{X} = \mathcal{G} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3$$

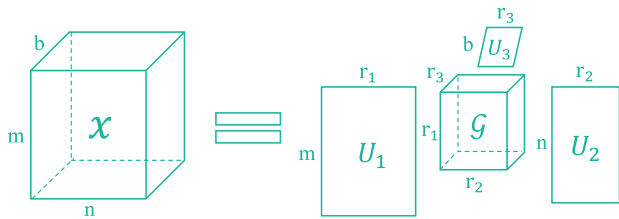$$\|\mathcal{Y} - \mathcal{X} - \mathcal{S}\|_F^2 \leq \varepsilon$$

Fig. 3.   Illustration of 3-D tensor Tucker decomposition.

where factorization matrices $\mathbf{U}_1 \in R^{m \times r_1}$, $\mathbf{U}_2 \in R^{n \times r_2}$, and $\mathbf{U}_3 \in R^{b \times r_3}$ are orthogonal in both spatial modes and spectral modes, respectively, and $\mathcal{G} \in R^{r_1 \times r_2 \times r_3}$ is called the core tensor. The 3-D tensor Tucker decomposition is intuitively shown in Fig. 3.

D. Total Variation-Regularized Low-Rank Matrix/Tensor Restoration of HSI

As reported in [29], each HSI band can be considered as a gray-level image and thus it has local piecewise smoothness in the spatial dimension. Therefore, TV regularization, first proposed in [30], has been widely introduced to the low-rank matrix/tensor decomposition framework to simultaneously exploit the spatial and spectral prior information. The related TV-regularized low-rank model can be formulated as follows:

$$\min_{\mathcal{X},\mathcal{S}} J_1(\mathcal{X}) + \lambda_1 \|\mathcal{X}\|_{TV} + \lambda_2 \|\mathcal{S}\|_1 \text{ s.t. } |\mathcal{Y} - \mathcal{X} - \mathcal{S}\|_F^2 \leq \varepsilon$$

where $J_1(\mathcal{X})$ is the different low-rank matrix/tensor HSI approximation and $\|\mathcal{X}\|_{TV}$ is the HSI TV regularization. Until now, various TV types, including band-by-band TV [29]; spatial–spectral TV [34], [39], [40]; and 3-D spectral–spatial cross TV [42], have been introduced to describe the term $\|\mathcal{X}\|_{TV}$. Among them, anisotropic spatial–spectral TV regularization with the low-rank Tucker decomposition method [40] has achieved the state-of-the-art HSI restoration results.

The combination of low-rank tensor decomposition with spatial–spectral TV regularization [40] achieves satisfactory HSI restoration results because it can simultaneously consider the HSI global spatial–spectral correlation and spatial–spectral smoothness. However, it still poses several problems. Band-by-band TV and spatial–spectral TV-regularized-based models make full use of the sparse prior of the spatial–spectral difference images, which usually use the convex ℓ_1-norm to depict the sparse prior. Generally speaking, sparsity is a helpful constraint to promote the piecewise smooth structure of each band. However, sparse prior only characterizes the number of nonzero elements, ignoring the local group structure of the nonzero elements. Thus, we remedied this deficiency to better depict the prior of HSI and improve HSI restoration results.

III. PROPOSED GROUP SPARSITY-REGULARIZED LOW-RANK TENSOR DECOMPOSITION MODEL

A. Group Sparsity Regularization

The mixed-norm regularizer has been widely used to exploit the shared sparse pattern of HSIs [45], [46]. Fig. 2 shows

the sparse pattern and shared sparse pattern. The gray squares represent the zero elements and the red squares represent the nonzero elements in the image. From Fig. 2(a) and (b), we can characterize both of them with sparsity regularization, because the number of zero elements is larger than that of the nonzero elements. However, the sparse pattern shown in Fig. 2(b) is different from that of Fig. 2(a). In Fig. 2(a), the nonzero elements are randomly distributed. In contrast, the nonzero elements shown in Fig. 2(b) present a shared row sparse pattern. This shared sparse pattern is also known as group sparsity and has been efficiently exploited by the convex mixed $\ell_{2,1}$-norm [45]. Group sparsity has also proved to be useful in HSI processing and application. For example, Iordache et al. [45] adopted group sparsity to regularize the abundance of nearby pixels to boost the unmixing performance. Chen et al. [47] introduced group sparsity regularization to exploit the sparsity of stripe components and achieved better destriping results than that using sparsity regularization [48].

For HSI, the imaging scene of different bands is the same. Band-by-band TV and spatial–spectral TV-regularized models employ the ℓ_1-norm to constrain the difference image of each band and explore the sparsity of the difference image from both spatial and spectral perspectives. However, given that the imaging scene of different bands is the same, the piecewise smooth structure of different bands should also be the same. That is to say, the difference images of the entire bands should obey the group sparse structure along the spectral dimension.

As shown in Fig. 1, the shared group sparse pattern of the spatial difference image occurs along the spectral dimension. Fig. 1(a) and (b) shows the spatial difference image along the spectral dimension for a 191-band 256×256 pixel after reshaping the 2-D difference image to a vector. From Fig. 1(a) and (b), it can be observed that the zero elements and nonzero elements approximately present the shared row pattern along the spectral dimension, and the structure of the visualization is similar to the group sparse pattern shown in Fig. 2(b). Moreover, we count the histogram distribution of the ℓ_2-norm value of all difference spectral vectors. From the histogram distribution shown in Fig. 2(c) and (d), most of the ℓ_2-norm values tend to be zero, indicating that most rows tend to be all zeros. Based on this, we introduce a group sparse regularization to constrain the spatial difference image in which each spectral signature is arranged as a group. Thus, the regularization term for $J_1(\mathcal{X})$ can be modeled as follows:

$$J_1(\mathcal{X}) = \|D\mathcal{X}\|_{2,1} \tag{3}$$

where D is the two differential operators in the two spatial dimensions D_x and D_y. Let $\mathcal{X}(i, j, k)$ is the pixel value at (i, j, k) in the HSI; then, the two differential operators are calculated as follows:

$$\begin{cases} D_x\mathcal{X}(i, j, k) = \mathcal{X}(i, j+1, k) - \mathcal{X}(i, j, k) \\ D_y\mathcal{X}(i, j, k) = \mathcal{X}(i+1, j, k) - \mathcal{X}(i, j, k). \end{cases}$$

To promote group sparsity, we utilized a weighted group sparse regularization to improve the group sparsity of the spatial difference image [49]. Thus, the regularization term (3) can be rewritten as follows:

$$J_1(\mathcal{X}) = \|W \odot D\mathcal{X}\|_{2,1}$$

and the weighted $\ell_{2,1}$-norm of $D\mathcal{X}$ is formulated as

$$\|W \odot D\mathcal{X}\|_{2,1} = \sum_{i=1}^{m}\sum_{j=1}^{n} W_x(i,j)\|D_x\mathcal{X}(i,j,:)\|_2$$
$$+ \sum_{i=1}^{m}\sum_{j=1}^{n} W_y(i,j)\|D_y\mathcal{X}(i,j,:)\|_2.$$

B. Group Sparsity-Regularized Low-Rank Tensor Decomposition Model

As in model (2), the key problem is to exploit the prior knowledge and design the corresponding regularization term for a clean HSI $\mathcal{X}$. Based on the aforementioned analysis, we found that the global spatial–spectral correlation of HSI can be efficiently formulated using low-rank Tucker decomposition. Moreover, we showed that the group sparse regularization of the spatial difference image is more reasonable and efficient than the popular TV regularization. Combining the Tucker decomposition and the group sparsity regularization, we propose a weighted group sparsity-regularized low-rank tensor decomposition (LRTDGS) for HSI restoration as follows:

$$\min_{\mathcal{X},\mathcal{S},\mathcal{G},\mathbf{U}_i} \quad \lambda_1\|W \odot D\mathcal{X}\|_{2,1} + \lambda_2\|\mathcal{S}\|_1$$
$$\text{s.t.} \quad \mathcal{X} = \mathcal{G} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3, \ \|\mathcal{Y}-\mathcal{X}-\mathcal{S}\|_F^2 \leq \varepsilon. \tag{4}$$

In our LRTDGS model, the priors of the clean HSI $\mathcal{X}$ and sparse noise $\mathcal{S}$ are fully exploited. The low-rank Tucker decomposition can maintain the spatial and spectral correlations and the weighted group sparse constraint is utilized to preserve the shared local smoothness in all bands. The ℓ_1-norm is employed to isolate the sparse noise, and the Frobenius norm is used to eliminate the Gaussian noise. Using an alternate updating process, we gradually isolated the sparse noise, removed the Gaussian noise from the noisy image, and separated the clean image. In the next section, we introduce an efficient iterative algorithm to optimize the proposed LRTDGS model.

C. Optimization

It is difficult to directly solve the variables $\mathcal{X}$ and $\mathcal{S}$ from model (4). A popular algorithm based on ALM [50] was utilized to efficiently optimize our HSI restoration model LRTDGS. In the following text, we present the ALM optimization steps in detail.

By introducing two auxiliary variables $\mathcal{R}$ and $\mathcal{Q}$, it is clear that model (4) is equivalent to the following:

$$\min_{\mathcal{X},\mathcal{S},\mathcal{G},\mathbf{U}_i,\mathcal{R},\mathcal{Q}} \quad \lambda_1\|W \odot \mathcal{R}\|_{2,1} + \lambda_2\|\mathcal{S}\|_1$$
$$\text{s.t.} \quad \mathcal{X} = \mathcal{G} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3, \ \|\mathcal{Y}-\mathcal{X}-\mathcal{S}\|_F^2$$
$$< \varepsilon, \ \mathcal{X} = \mathcal{Q}, \ D\mathcal{Q} = \mathcal{R}.$$

Based on the ALM algorithm, we transformed the constraint problem into an optimization of the following augmented Lagrangian function:

$$L_\beta(\mathcal{X},\mathcal{S},\mathcal{G},\mathbf{U}_i,\mathcal{R},\mathcal{Q},\mathcal{W}_j)$$
$$= \lambda_1\|W \odot \mathcal{R}\|_{2,1} + \lambda_2\|\mathcal{S}\|_1$$

$$+ \ <\mathcal{W}_1, \mathcal{Y}-\mathcal{X}-\mathcal{S}> +\frac{\beta}{2}\|\mathcal{Y}-\mathcal{X}-\mathcal{S}\|_F^2+ <\mathcal{W}_2, \mathcal{X}$$
$$- \ \mathcal{Q}> +\frac{\beta}{2}\|\mathcal{X}-\mathcal{Q}\|_F^2+ <\mathcal{W}_3, D\mathcal{Q}-\mathcal{R}>$$
$$+ \ \frac{\beta}{2}\|D\mathcal{Q}-\mathcal{R}\|_F^2$$

under the constraint of low-rank Tucker decomposition $\mathcal{X} = \mathcal{G} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3$, $\mathbf{U}_i^T\mathbf{U}_i = \mathbf{I}(i = 1, 2, 3)$, where $<\cdot,\cdot>$ is the inner product of the two tensors; β is a positive penalty parameter; and $\mathcal{W}_j(j = 1, 2, 3)$ are the Lagrange multipliers. In the ALM framework, we alternately optimized one variable while fixing the other variables in an iterative manner. Thus, we transformed the difficult multivariable optimization problem into more easily solvable subproblems. In the next section, we present how each of these subproblems can be solved.

1) The $\mathcal{R}$ subproblem is provided as follows:

$$\min_{\mathcal{R}} \quad \lambda_1\|W \odot \mathcal{R}\|_{2,1}+ <\mathcal{W}_3, D\mathcal{Q}-\mathcal{R}>$$
$$+ \frac{\beta}{2}\|D\mathcal{Q}-\mathcal{R}\|_F^2$$
$$= \arg\min_{\mathcal{R}} \quad \lambda_1\|W \odot \mathcal{R}\|_{2,1} + \frac{\beta}{2}\left\|D\mathcal{Q}-\mathcal{R}+\frac{\mathcal{W}_3}{\beta}\right\|_F^2.$$

Let $D\mathcal{Q} + (\mathcal{W}_3/\beta) = \mathcal{C}$, the closed-form solution of the tube fibers of $\mathcal{R}$ is calculated by (see [51]) the following:

$$\mathcal{R}(i,j,:) = \begin{cases} \frac{\|\mathcal{C}(i,j,:)\|_2-\frac{W(i,j)\lambda_1}{\beta}}{\|\mathcal{C}(i,j,:)\|_2}\mathcal{C}(i,j,:), & \text{if } \frac{W(i,j)\lambda_1}{\beta} \\ & < \|\mathcal{C}(i,j,:)\|_2 \\ \mathbf{0}, & \text{otherwise.} \end{cases} \tag{5}$$

2) $\mathcal{G}$, $\mathbf{U}_i$, and $\mathcal{X}$ Subproblems: By fixing the other variables, the subproblems are as follows:

$$\min_{\mathcal{X}=\mathcal{G}\times_1\mathbf{U}_1\times_2\mathbf{U}_2\times_3\mathbf{U}_3} \quad <\mathcal{W}_1, \mathcal{Y}-\mathcal{X}-\mathcal{S}>$$
$$+\frac{\beta}{2}\|\mathcal{Y}-\mathcal{X}-\mathcal{S}\|_F^2+ <\mathcal{W}_2, \mathcal{X}-\mathcal{Q}> +\frac{\beta}{2}\|\mathcal{X}-\mathcal{Q}\|_F^2.$$

We equivalently transformed the equation as

$$\arg\min_{\mathcal{G},\mathbf{U}_i} \quad \beta\left\|\mathcal{G} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3 \right.$$
$$\left. -\frac{1}{2}\left(\mathcal{Y}-\mathcal{S}+\mathcal{Q}+\frac{\mathcal{W}_1-\mathcal{W}_2}{\beta}\right)\right\|_F^2. \tag{6}$$

The Tucker decomposition factors $\mathcal{G}$ and $\mathbf{U}_i$ can be efficiently solved by using the higher-order orthogonal iteration (HOOI) algorithm [44]. When we obtained the decomposition factors $\mathcal{G}$ and $\mathbf{U}_i$, we then updated $\mathcal{X}$ as follows:

$$\mathcal{X} = \mathcal{G} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3. \tag{7}$$

3) The $\mathcal{Q}$ subproblem is optimized by minimizing the following problem:

$$\arg\min_{\mathcal{Q}} \quad \frac{\beta}{2}\left\|\mathcal{X}-\mathcal{Q}+\frac{\mathcal{W}_2}{\beta}\right\|_F^2 + \frac{\beta}{2}\left\|D\mathcal{Q}-\mathcal{R}+\frac{\mathcal{W}_3}{\beta}\right\|_F^2$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

6   IEEE TRANSACTIONS ON CYBERNETICS

which is a least squared problem and can be optimized via the following linear equations system:

$$\left(\mathcal{I} + D^T D\right)\mathcal{Q} = \left(\mathcal{X} + \frac{\mathcal{W}_2}{\beta}\right) + D^T\left(\mathcal{R} - \frac{\mathcal{W}_3}{\beta}\right)$$

where $\mathcal{I}$ denotes the unit tensor. In this article, the periodic boundary condition was considered for $\mathcal{Q}$, then $D^T D$ possesses block circulant matrices with a circulant block structure. The efficient fast Fourier transform (FFT) can be employed to solve this equation as follows:

$$\mathcal{Q} = \text{ifftn}\left(\frac{\mathcal{B}}{\beta\mathbf{1} + \beta|\text{fftn}(D)|^2}\right) \quad (8)$$

where $\mathcal{B} = \text{fftn}((\mathcal{X} + (\mathcal{W}_2/\beta)) + D^T(\mathcal{R} - (\mathcal{W}_3/\beta)))$; and fftn and ifftn are the fast 3-D Fourier transform and its inverse transform. The division is also performed elementwisely and $|\cdot|^2$ is the elementwise square.

4) $\mathcal{S}$ subproblem has the following form:

$$\min_{\mathcal{S}} \ \lambda_2\|\mathcal{S}\|_1 + \langle \mathcal{W}_1, \mathcal{Y} - \mathcal{X} - \mathcal{S}\rangle + \frac{\beta}{2}\|\mathcal{Y} - \mathcal{X} - \mathcal{S}\|_F^2$$
$$= \arg\min_{\mathcal{S}} \lambda_2\|\mathcal{S}\|_1 + \frac{\beta}{2}\left\|\mathcal{Y} - \mathcal{X} - \mathcal{S} + \frac{\mathcal{W}_1}{\beta}\right\|_F^2.$$

By applying a soft-threshold shrinkage operator, the $\mathcal{S}$ subproblem can be accurately solved using the following:

$$\mathcal{S} = \textbf{Shrinkage}\left(\mathcal{Y} - \mathcal{X} + \frac{\mathcal{W}_1}{\beta}, \frac{\lambda_2}{\beta}\right) \quad (9)$$

where $\textbf{Shrinkage}(a, b) = \text{sign}(a)\max(|a| - b, 0)$.

5) *Multiplier Updating:* According to the ALM algorithm, we updated the Lagrange multipliers $\mathcal{W}_i(i = 1, 2, 3)$ as follows:

$$\begin{cases} \mathcal{W}_1 = \mathcal{W}_1 + \beta(\mathcal{Y} - \mathcal{X} - \mathcal{S}) \\ \mathcal{W}_2 = \mathcal{W}_2 + \beta(\mathcal{X} - \mathcal{Q}) \\ \mathcal{W}_3 = \mathcal{W}_3 + \beta(D\mathcal{Q} - R). \end{cases} \quad (10)$$

Moreover, the weight W used for the next iteration was updated from the current solution of the formulation (5) [49]

$$W(i, j) = \frac{1}{\left\|(D\mathcal{Q} + \frac{\mathcal{W}_3}{\beta})(i, j, :)\right\|_2 + \text{eps}}$$

where eps is a small number used to avoid singularities. 1)–5) processes indicate one ALM iteration that transforms the difficult multivariable optimization problem to more easily solvable subproblems [i.e., 1)–5) subproblems]. Summarizing the processes of steps 1)–5), we obtained the ALM algorithm to solve the proposed HSI restoration model LRTDGS (4), as shown in Algorithm 1.

D. Analysis of Computational Complexity

Suppose the size of the HSI input is $m \times n \times b$, we present the computational complexity of the LRTDGS solution. The updating of $\mathcal{R}$ and $\mathcal{S}$ subproblems is soft threshold shrinkage operations; then, the total computational complexity of $\mathcal{R}$ and $\mathcal{S}$ solvers is $O(2mnb)$. The $\mathcal{X}$, $\mathcal{G}$, and $\mathbf{U}_i$ subproblems are updated using the HOOI algorithm, which is needed to implement the SVDs and tensor-matrix product operation.

Algorithm 1 Optimization Process for LRTDGS Solver

Input: Noisy image $\mathcal{Y}$, regularization parameters λ_1 and λ_2, estimated rank (r_1, r_2, r_3) for Tucker decomposition, ε, k_{max}, β_{max}, and ρ.

1: Initialize: Let $\mathcal{X} = \mathcal{Y}$, $\mathcal{Q} = \mathcal{R} = \mathcal{W}_1 = \mathcal{W}_2 = \mathcal{W}_3 = \mathbf{0}$, $k = 0$, and β
2: **while** not coverged **do**
3:    Compute $\mathcal{R}$ via (5).
4:    Update $\mathcal{G}$, $\mathbf{U}_i$ and $\mathcal{X}$ via (6) and (7).
5:    Compute $\mathcal{Q}$ via FFT (8).
6:    Update $\mathcal{S}$ via (9).
7:    Compute the Lagrange multipliers $\mathcal{W}_1$, $\mathcal{W}_2$ and $\mathcal{W}_3$ by (10).
8:    Update the penalty parameter $\beta = \min\{\rho\beta, \beta_{max}\}$.
9:    Check the convergence condition $\frac{\|\mathcal{X}^{k+1} - \mathcal{X}^k\|_F}{\|\mathcal{X}^k\|_F} \leq \varepsilon$ and $k < k_{max}$.
10: **end while**

Output: The restoration result $\mathcal{X}$.

For simplified analysis, we let $r = r_1 = r_2 = r_3$; then, the total computational complexity of HOOI is $O(mnbr + (m + n + b)r^4 + r^6)$ [52]. Moreover, FFT was employed to optimize the $\mathcal{Q}$ subproblem, which needs $O(mnb \log(mnb))$ calculations. Considering all subproblems, the overall computational complexity of each iteration in algorithm 1 is $O(2mnb + mnbr + (m + n + b)r^4 + r^6 + mnb \log(mnb))$.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we will present extensive experiments based on simulated and real datasets to evaluate the performances of the proposed LRTDGS method. To better demonstrate the superiority of the combination of low-rank tensor decomposition and group sparsity regularization, we compared the performance of our method to five other state-of-the-art HSI restoration approaches. These approaches include LRMR [4], the combination of low-rank matrix factorization with TV regularization (LRTV) [29], the spatiospectral TV model (SSTV)[1] [35], weighted low-rank constraint with the spatial–spectral TV regularization model (LSSTV) [36], and anisotropic spatial–spectral TV regularization with low-rank tensor decomposition (LRTDTV)[2] [40]. For the parameter selection of these methods in all experiments, we carefully followed the authors' suggestions in their papers to guarantee optimal results. Before the experiments, the pixel values of each band were scaled within the range of [0, 1]. All the experiments were conducted in MATLAB R2018b using a laptop of 32-GB RAM, with a Intel Core i7-8750H CPU@2.20 GHz.

A. Simulated Experiments

1) Experimental Setting: To show the robustness of the LRTDGS method, we selected two clean HSI datasets for the simulated experiments. The first clean HSI was the simulated Indian Pines dataset, which was also used in [29], [36],

[1] https://sites.google.com/view/hkaggarwal/publications
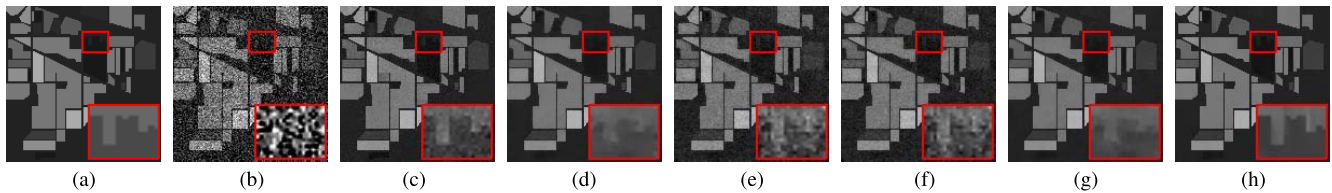[2] http://gr.xjtu.edu.cn/web/dymeng/3

Fig. 4. Restoration results of all the comparison methods using band 224 of the Indian Pines data under case 1. (a) Original. (b) Noisy. (c) LRMR. (d) LRTV. (e) SSTV. (f) LSSTV. (g) LRTDTV. (h) LRTDGS.
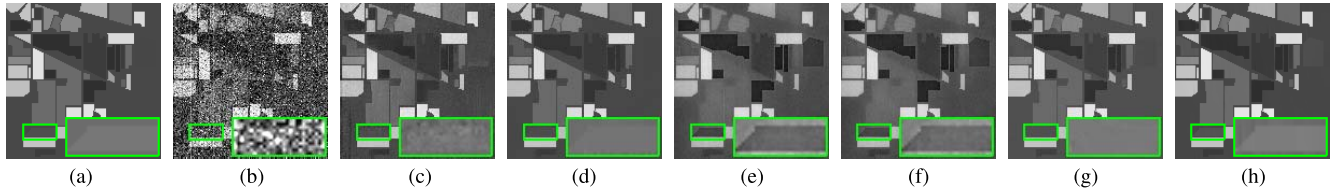

Fig. 5. Restoration results of all comparison methods using band 118 of the Indian Pines data under case 6. (a) Original. (b) Noisy. (c) LRMR. (d) LRTV. (e) SSTV. (f) LSSTV. (g) LRTDTV. (h) LRTDGS.

and [40]. This data contains 145×145 pixels in the spatial domain and 224 spectral bands. The HYDICE Washington DC Mall data adopted in [4] and [35] was chosen as the second dataset, with a size of $256 \times 256 \times 191$. To simulate complicated noise cases in a real scene, six different types of noise cases were added to these two clean HSI datasets. A detailed description of these cases can be presented as follows.

Case 1: The zero-mean Gaussian noise was added to all bands, and the noise variance of each band was equal to 0.15.

Case 2: Different Gaussian noise variances with zero-means were added to each band and the noise variances were randomly selected within the range of [0, 0.2].

Case 3: Mixtures of Gaussian with impulse noises were added to each band and the Gaussian noise was added in the same manner as in case 2. Moreover, we added impulse noise in different percentages to each band and the percentages were randomly selected within the range of [0, 0.2].

Case 4: Mixtures of Gaussian with deadline noises were added to the HSI data. On the basis of case 2, we chose 40% of all bands and add deadline noise whose number was randomly selected within the range of 3–10.

Case 5: Mixtures of Gaussian with stripe noises were added to the HSI data. We added Gaussian noise in the same manner as in case 2 to each band, and 40% of all bands were selected in which to add stripe noise whose number was also randomly selected from 3 to 10.

Case 6: Mixtures of Gaussian, impulse, deadline, and stripe noises were added to the HSI data. Gaussian noise and impulse noise were added just as in case 3. In addition, we severally chose 20% of all bands in which to add deadline and stripe noises whose numbers were randomly selected from 3 to 10.

In the following text, we evaluated the HSI restoration performances of all comparison approaches from three perspectives, that is, visual, quantitative, and qualitative comparisons.

2) Visual Comparison: For a visual comparison, we chose two representative cases 1 and 6, to compare the performance of the different methods. Figs. 4 and 5 show the restoration results for bands 224 and 118 of the Indian Pines data

under these two cases, respectively. For a better visual comparison, we enlarged some areas in the image. As shown in Figs. 4 and 5, LRMR, SSTV, and LSSTV cannot completely eliminate the noise in the presented results, clearly observed in the enlarged box. LRTV and LRTDTV provide satisfactory restoration results compared to those of the other comparison methods, completely removing all noises. However, the image details cannot be preserved well shown in the enlarged box of Figs. 4 and 5. Compared with the other methods as shown in Figs. 4(h) and 5(h), it can be clearly seen that all of the mixed noises are eliminated and the image edges and details are effectively preserved in the restoration results, indicating the superiority of the LRTDGS method for HSI restoration in the Indian Pines dataset. The restoration performances of cases 1 and 6 for the Washington DC Mall data are shown in Figs. 6 and 7, respectively. Similar to the Indian Pines dataset, the LRMR, SSTV, and LSSTV results still contain some noises. LRTV, LRTDTV, and the proposed LRTDGS achieve similar visual results in these two cases. However, the image details are blurred by LRTDTV as shown in the enlarged boxes of Figs. 6(g) and 7(g). In comparison, our LRTDGS method simultaneously removes all noises and preserves the most original details in the image. In summary, our method achieves an obvious visual improvement for HSI restoration using the Indian Pines dataset, the reason being that these data possess a same local smoothness characteristic in all bands. Moreover, the proposed method can also achieve a competitive result compared with that of the other methods using the Washington DC Mall dataset.

3) Quantitative Comparison: The aforementioned visual comparison showed the effectiveness of our LRTDGS method. In the following text, we describe employing five objective quantitative evaluation indices to demonstrate the performance of the proposed LRTDGS method in all simulated experiments. These evaluation indices are the peak signal-to-noise ratio (PSNR), structure similarity (SSIM) [53], feature similarity (FSIM) [54], erreur relative globale adimensionnelle de synthese (ERGAS), and spectral angle mapper (SAM).
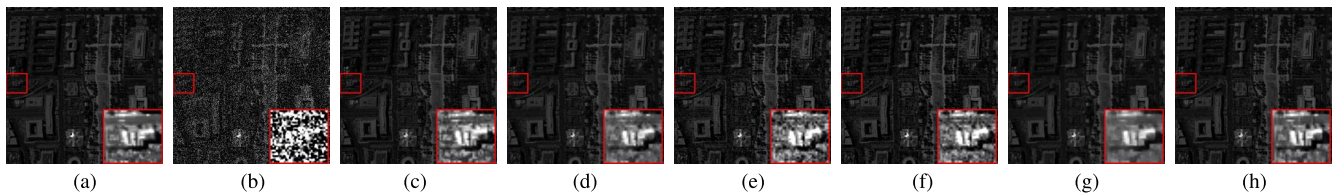
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

8   IEEE TRANSACTIONS ON CYBERNETICS

Fig. 6.   Restoration results of all the comparison methods using band 161 of the Washington DC Mall data under case 1. (a) Original. (b) Noisy. (c) LRMR. (d) LRTV. (e) SSTV. (f) LSSTV. (g) LRTDTV. (h) LRTDGS.
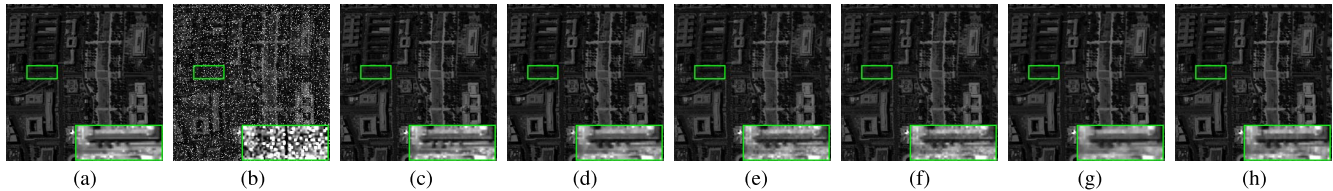

Fig. 7.   Restoration results of all the comparison methods using band 155 of the Washington DC Mall data under case 6. (a) Original. (b) Noisy. (c) LRMR. (d) LRTV. (e) SSTV. (f) LSSTV. (g) LRTDTV. (h) LRTDGS.
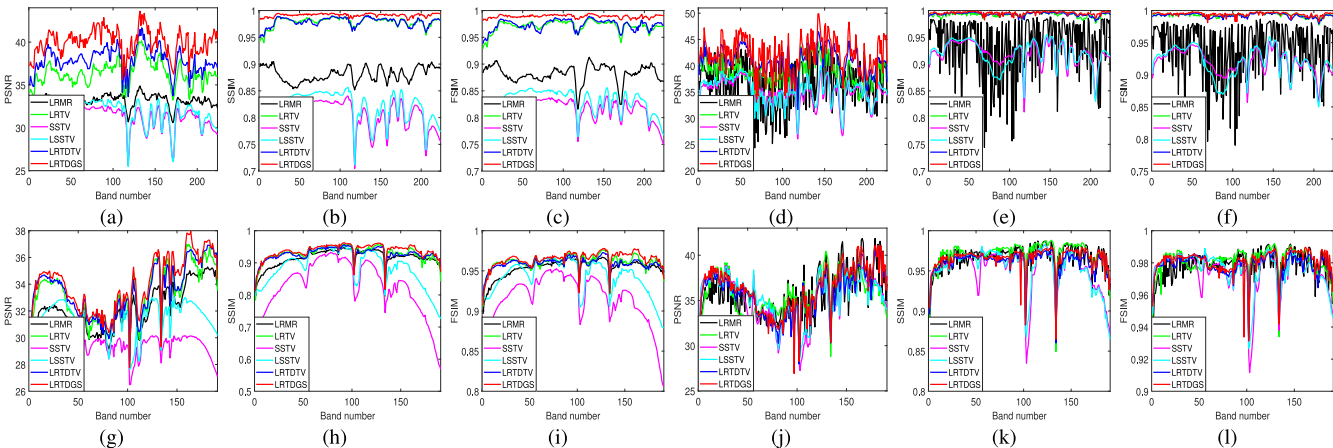

Fig. 8.   PSNR, SSIM, and FSIM values of all bands of the different comparison approaches in cases 1 and 6. (a)–(c) Indian Pines dataset for case 1. (d)–(f) Indian Pines dataset for case 6. (g)–(i) Washington DC Mall for case 1. (j)–(l) Washington DC Mall for case 6.

In our experiments, the PSNR, SSIM, and FSIM results are the average values of those for all bands. The higher values indicate better restoration results for the PSNR, SSIM, and FSIM evaluation indices. As for the ERGAS and SAM indices, the lower values indicate better quality restoration results.

Table I presents the five quantitative evaluation indices of all the comparison methods under the aforementioned six different noise cases for the Indian Pines and the Washington DC Mall datasets. The highlighted bold value of each index is the best result among all the comparison methods. From the results, because of the combination of the low-rank tensor decomposition with spatial–spectral TV regularization, LRTDTV obtained better results than those of the other matrix-based format methods due to the fact that low-rank tensor decomposition can capture the global correlation in the spatial–spectral dimensions. Compared with the matrix-based format methods and LRTDTV, the proposed LRTDGS achieves the best performance for all evaluation indices under nearly all cases. This is because the traditional TV regularization fails to explore the global structure of the difference images, but our model remedies this deficiency. Particularly,

our method achieves an improvement in all metrics for the Indian Pines dataset. In the Washington DC Mall dataset, the proposed method can also achieve better results compared to those of all comparison methods. This is because the Indian Pines dataset contains the same local smoothness characteristic in all bands, which can be efficiently captured using group sparsity to constrain the difference images. In summary, the quantitative comparison of the six different noise cases for the two clean datesets shows the effectiveness of our LRTDGS method.

4) Qualitative Comparison: In the aforementioned discussion, we presented the average value of the evaluation indices to evaluate the performance. Fig. 8 shows the curves of PSNR, SSIM, and FSIM evaluation indices for each band based on cases 1 and 6. As presented in Fig. 8(a)–(f), the proposed LRTDGS method achieves the highest PSNR, SSIM, and FSIM values in all bands of the Indian Pines dataset. This may because the group sparsity constraint of the difference image can better preserve the most details in each band. As shown in Fig. 8(j)–(l), our method did not achieve the best evaluation in every each band, but we achieved the average optimum in all bands.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

CHEN *et al.*: HYPERSPECTRAL IMAGE RESTORATION USING WEIGHTED GROUP SPARSITY-REGULARIZED LOW-RANK TENSOR DECOMPOSITION　　9

TABLE I
QUANTITATIVE COMPARISON OF ALL COMPARISON METHODS UNDER THE AFOREMENTIONED SIX DIFFERENT
NOISE CASES FOR THE INDIAN PINES AND WASHINGTON DC MALL DATASETS

Dataset	Case	Index	Noisy	LRMR	LRTV	SSTV	LSSTV	LRTDTV	LRTDGS
Indian Pines	Case 1	PSNR	16.474	33.092	36.322	31.183	31.763	37.876	**40.009**
		SSIM	0.2632	0.8805	0.9790	0.8011	0.8169	0.9791	**0.9915**
		FSIM	0.3830	0.8816	0.9711	0.8146	0.8290	0.9733	**0.9891**
		ERGAS	351.057	52.899	37.271	68.310	64.721	31.497	**24.602**
		SAM	0.2915	0.0359	0.0239	0.0510	0.0477	0.0207	**0.0155**
	Case 2	PSNR	23.742	38.794	40.244	36.298	37.786	42.710	**44.176**
		SSIM	0.4699	0.9506	0.9939	0.9372	0.9517	0.9954	**0.9973**
		FSIM	0.5551	0.9552	0.9938	0.9371	0.9507	0.9949	**0.9973**
		ERGAS	269.259	33.159	25.475	41.759	35.171	19.385	**16.778**
		SAM	0.2277	0.0266	0.0156	0.0297	0.0254	0.0125	**0.0103**
	Case 3	PSNR	14.741	37.487	39.170	35.427	35.918	41.480	**42.856**
		SSIM	0.2387	0.9393	0.9924	0.9240	0.9257	0.9938	**0.9962**
		FSIM	0.4241	0.9465	0.9930	0.9246	0.9252	0.9930	**0.9958**
		ERGAS	467.590	37.883	28.200	45.654	42.289	22.157	**19.400**
		SAM	0.3760	0.0300	0.0168	0.0326	0.0310	0.0138	**0.0129**
	Case 4	PSNR	20.906	36.044	38.982	35.671	36.235	41.987	**43.169**
		SSIM	0.4438	0.9387	0.9928	0.9340	0.9372	0.9950	**0.9968**
		FSIM	0.5388	0.9448	0.9937	0.9340	0.9365	0.9944	**0.9966**
		ERGAS	310.882	50.801	28.899	44.175	40.765	20.478	**18.863**
		SAM	0.2662	0.0383	0.0182	0.0324	0.0308	0.0133	**0.0132**
	Case 5	PSNR	22.479	38.362	40.048	36.069	37.439	42.511	**43.932**
		SSIM	0.4569	0.9494	0.9938	0.9356	0.9493	0.9951	**0.9972**
		FSIM	0.5467	0.9534	0.9936	0.9347	0.9475	0.9944	**0.9970**
		ERGAS	273.237	34.445	25.776	42.680	36.259	19.670	**17.139**
		SAM	0.2311	0.0277	0.0163	0.0305	0.0264	0.0126	**0.0110**
	Case 6	PSNR	14.383	36.506	38.804	34.845	35.211	41.004	**42.166**
		SSIM	0.2333	0.9352	0.9921	0.9196	0.9195	0.9933	**0.9954**
		FSIM	0.4174	0.9406	0.9928	0.9199	0.9192	0.9923	**0.9944**
		ERGAS	482.377	44.689	29.101	47.999	45.483	23.197	**21.189**
		SAM	0.3914	0.0361	0.0180	0.0354	0.0347	**0.0151**	0.0152
Washington DC Mall	Case 1	PSNR	16.479	32.144	33.228	29.342	31.491	33.459	**33.961**
		SSIM	0.2800	0.9158	0.9283	0.8268	0.8859	0.9275	**0.9369**
		FSIM	0.6334	0.9540	0.9610	0.9153	0.9426	0.9592	**0.9645**
		ERGAS	633.769	101.804	90.270	145.804	113.069	86.811	**82.748**
		SAM	0.7345	0.1643	0.1384	0.2398	0.1873	**0.1158**	0.1269
	Case 2	PSNR	23.856	36.826	36.484	36.230	37.070	36.874	**37.541**
		SSIM	0.5264	0.9699	0.9685	0.9604	0.9660	0.9690	**0.9727**
		FSIM	0.7658	0.9815	0.9789	0.9779	0.9810	0.9830	**0.9847**
		ERGAS	472.944	62.532	63.300	67.283	61.397	59.518	**56.191**
		SAM	0.6217	0.1086	**0.0855**	0.1254	0.1161	0.0860	0.0920
	Case 3	PSNR	13.840	36.056	35.740	35.326	35.593	35.907	**36.462**
		SSIM	0.2347	0.9666	**0.9676**	0.9523	0.9542	0.9616	0.9658
		FSIM	0.6273	0.9793	0.9792	0.9747	0.9762	0.9788	**0.9810**
		ERGAS	775.179	67.541	91.656	73.809	71.005	66.341	**63.769**
		SAM	0.7902	0.1101	0.1567	0.1254	0.1209	**0.0910**	0.0945
	Case 4	PSNR	22.537	36.547	36.352	36.039	36.769	36.748	**37.381**
		SSIM	0.5158	0.9698	0.9683	0.9596	0.9650	0.9685	**0.9723**
		FSIM	0.7608	0.9813	0.9788	0.9776	0.9806	0.9827	**0.9845**
		ERGAS	481.408	64.337	63.976	68.535	63.262	60.474	**57.288**
		SAM	0.6368	0.1094	0.0861	0.1298	0.1219	**0.0877**	0.0938
	Case 5	PSNR	22.967	36.656	36.311	36.131	36.944	36.831	**37.390**
		SSIM	0.5204	0.9691	0.9714	0.9598	0.9654	0.9687	**0.9719**
		FSIM	0.7624	0.9811	0.9812	0.9777	0.9806	0.9829	**0.9841**
		ERGAS	476.975	63.602	81.357	68.017	62.220	59.735	**57.280**
		SAM	0.6259	0.1103	0.1456	0.1276	0.1186	**0.0860**	0.0932
	Case 6	PSNR	13.786	35.759	35.542	35.152	35.406	35.794	**36.278**
		SSIM	0.2311	0.9655	**0.9663**	0.9513	0.9531	0.9612	0.9654
		FSIM	0.6245	0.9784	0.9781	0.9743	0.9757	0.9786	**0.9806**
		ERGAS	779.013	69.753	127.723	74.984	72.300	67.197	**64.168**
		SAM	0.7948	0.1128	0.1463	0.1286	0.1242	**0.0923**	0.0939

Furthermore, we present the spectral signatures of the restoration results compared with the original image to further report the qualitative comparison. To show the performance of the different methods in edge preservation, we selected an edge pixel of location (30, 74) and showed the spectral signatures between the restored image and the original
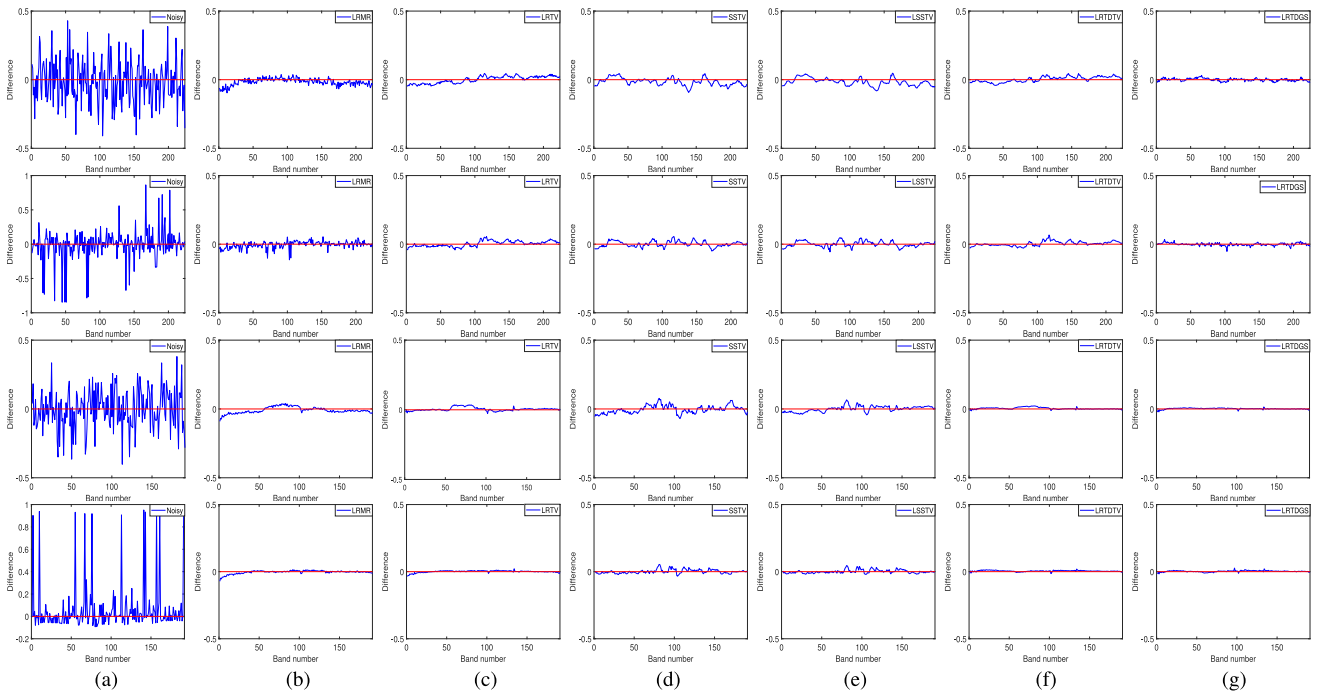
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Fig. 9. Difference between the original spectrum and the restoration results of pixel (30, 74) on the Indian Pines image and pixel (10, 162) on the Washington DC Mall image under cases 1 and 6. From top to bottom are the Indian Pines dataset for case 1, Indian Pines dataset for case 6, Washington DC Mall for case 1, and Washington DC Mall for case 6. (a) Noisy. (b) LRMR. (c) LRTV. (d) SSTV (e) LSSTV. (f) LRTDTV. (g) LRTDGS.
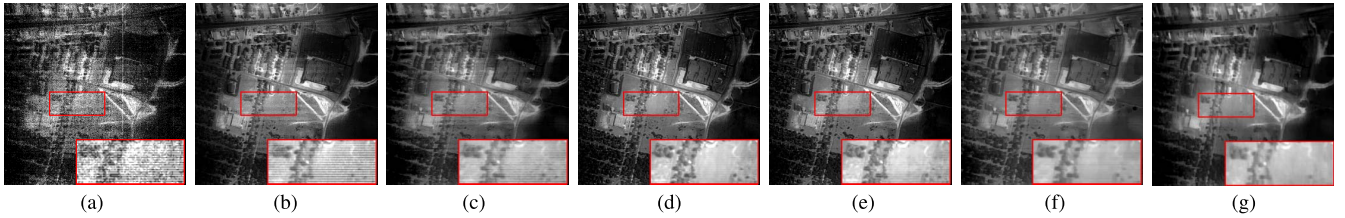

Fig. 10. Restoration results of all comparison methods for band 139 of the real Urban dataset. (a) Original. (b) LRMR. (c) LRTV. (d) SSTV. (e) LSSTV. (f) LRTDTV. (g) LRTDGS.

Indian Pines dataset. Furthermore, we selected a smooth pixel of location (10, 162) for illustration in the Washington DC Mall. Fig. 9 shows the spectral signature difference under noise cases 1 and 6. It can be clearly observed that the proposed LRTDGS method achieves the signatures most similar to the originals. LRTDTV also achieved better results in the comparison of the Washington DC Mall dataset, because the low-rank tensor decomposition can efficiently perverse the spatial–spectral correlation.

B. Real Experiments

In the simulated experiments, we showed the performance of our proposed LRTDGS method in a variety of noise cases. In the following text, two datasets were employed to illustrate the effectiveness of our method in the real-world. These two datasets are the Hyperspectral Digital Imagery Collection Experiment (HYDICE) urban dataset[3] and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)

Indian Pines dataset.[4] As a preprocessing, the pixel value of each band was normalized to [0, 1].

1) HYDICE Urban Dataset: The HYDICE Urban is a popular dataset and is extensively applied to HSI restoration experiments. The tensor size of this data is $307 \times 307 \times 210$ and is severely contaminated by the Gaussian noise, deadlines, stripe noise, water absorption, atmosphere, and some unknown noises. For succinctness, we selected two typical bands to present the restoration results of the different approaches. Figs. 10 and 11 show the restoration results of bands 139 and 207 in this data, respectively. From Figs. 10(a) and 11(a), one can see that these two bands are completely contaminated by a variety of noises, including Gaussian, stripe, and deadlines noises. After denoising using the different HSI restoration methods, the noise is obviously removed. As shown in Fig. 10, the LRMR, LRTV, SSTV, and LSSTV methods cannot eliminate the stripes in the results, as observed in the enlarged box on the image. LRTDTV can obtain a better visual result than that of the other methods, but still a few stripes existed in the image as shown in Fig. 10(f). Moreover, LRTDTV changes the

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

CHEN *et al.*: HYPERSPECTRAL IMAGE RESTORATION USING WEIGHTED GROUP SPARSITY-REGULARIZED LOW-RANK TENSOR DECOMPOSITION 11
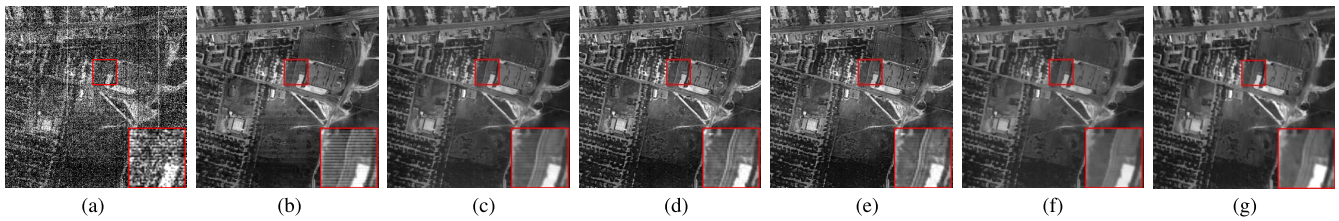

Fig. 11.   Restoration results of all comparison methods for band 207 of the real Urban dataset. (a) Original. (b) LRMR. (c) LRTV. (d) SSTV. (e) LSSTV. (f) LRTDTV. (g) LRTDGS.
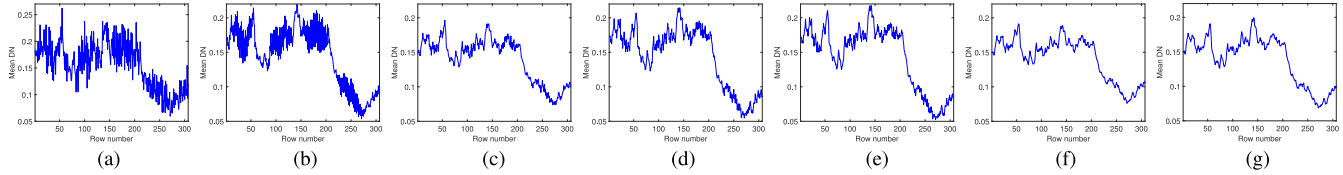

Fig. 12.   Horizontal mean profiles of band 207 for the real Urban dataset. (a) Original. (b) LRMR. (c) LRTV. (d) SSTV. (e) LSSTV. (f) LRTDTV. (g) LRTDGS.
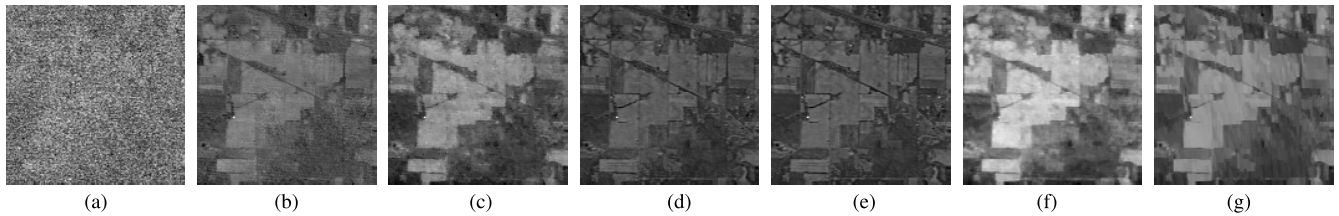

Fig. 13.   Restoration results of all comparison methods for band 150 of the real Indian Pines dataset. (a) Original. (b) LRMR. (c) LRTV. (d) SSTV. (e) LSSTV. (f) LRTDTV. (g) LRTDGS.
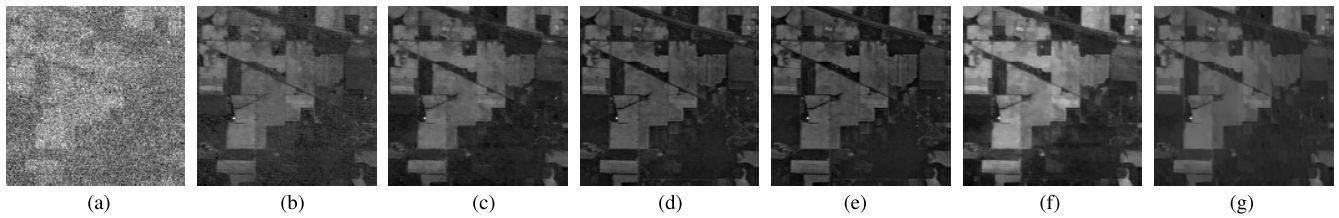

Fig. 14.   Restoration results of all comparison methods for band 220 of the real Indian Pines dataset. (a) Original. (b) LRMR. (c) LRTV. (d) SSTV. (e) LSSTV. (f) LRTDTV. (g) LRTDGS.

image contrast in the restoration result as shown in Fig. 10(f) because the left and right sides of the image are brightened compared to those of the other results. As shown in Fig. 11, the LRMR, LRTV, SSTV, and LSSTV methods also fail to eliminate the stripes. LRTDTV and our method can efficiently eliminate the stripes and other noises, but the LRTDTV result is smoothed in the image. By combining the low-rank tensor decomposition with group sparsity regularization, the proposed LRTDGS method completely removes all mixed noises and preserves the image detail as shown in Figs. 10 and 11(g).

To clearly illustrate the restoration results, the horizontal mean profiles of band 207 for the Urban dataset are shown in Fig. 12. The vertical axis denotes the mean digital number value of each row and the horizontal axis represents the row number. As shown in Fig. 12(a), many fluctuations occur in the curves because of the effects of stripes and other noises. It can be observed that the fluctuations are largely reduced using the different methods as shown in Fig. 12(b)–(g). As shown in Fig. 12(b)–(e), some minor fluctuations also remain in the curve, indicating the stripes in the image. This is

consistent with the restoration results shown in Fig. 11(b)–(e). Fig. 11(f) shows the smoothed result of LRTDTV, which can also be confirmed from Fig. 12(f). In comparison, the proposed method achieves more reasonable mean profile results as shown in Fig. 12(g).

2) AVIRIS Indian Pines Dataset: The second dataset is the AVIRIS Indian Pines with the size of $145 \times 145 \times 220$. Some bands in this dataset are seriously degraded by mixed Gaussian and impulse noises. Similar to the Urban data, we also chose two typical noise bands to display noise removal results. The restoration results of bands 150 and 220 in this dataset are shown in Figs. 13 and 14, respectively. As shown in Fig. 13(a), the band 150 image is severely contaminated by noise and the information of this band is completely corrupted. Fig. 14(a) shows that this band is also seriously corrupted by noise, and we cannot find only useful information. From the results, all methods can remove most of the noises and restore the image structure. However, the comparison methods cannot completely eliminate the noise, and the image details are destroyed, as shown in the visual results of Fig. 13.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

12   IEEE TRANSACTIONS ON CYBERNETICS
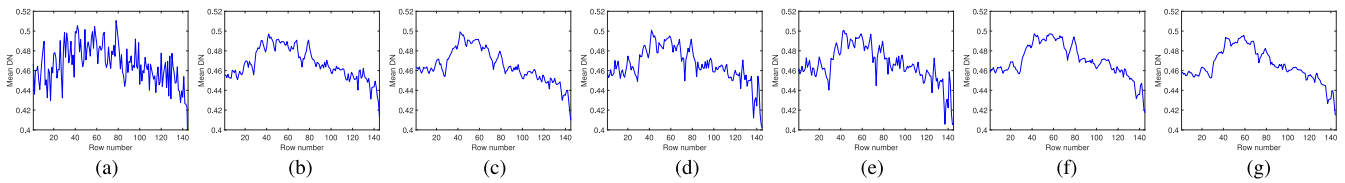

Fig. 15.   Horizontal mean profiles for band 150 using the real Indian Pines dataset. (a) Original. (b) LRMR. (c) LRTV. (d) SSTV. (e) LSSTV. (f) LRTDTV. (g) LRTDGS.

TABLE II
BLIND ASSESSMENT INDEX Q-METRIC COMPARISON ON THE REAL DATA

Dataset	Noisy	LRMR	LRTV	SSTV	LSSTV	LRTDTV	LRTDGS
Urban	0.063	0.072	0.069	0.066	0.066	0.071	**0.074**
Indian	0.064	0.074	0.074	0.075	0.076	0.079	**0.082**

TABLE III
QUANTITATIVE COMPARISON OF LRTDGS WITH AND
WITHOUT WEIGHT ON THE INDIAN PINES DATASET

Case	Method	PSNR	SSIM	FSIM	ERGAS	SAM
Case 1	LRTDW	36.262	0.9716	0.9588	37.676	0.0237
	LRTDGS	**40.009**	**0.9915**	**0.9891**	**24.602**	**0.0155**
Case 2	LRTDW	41.726	0.9912	0.9870	21.019	0.0138
	LRTDGS	**44.176**	**0.9973**	**0.9973**	**16.778**	**0.0103**

As shown in Fig. 14, the results of LRMR, LRTV, SSTV, and LSSTV methods still contain some noise and the image structures are destroyed. LRTDTV can efficiently remove the noise but it oversmooths the image and changes the contrast as shown in Fig. 14(f). Compared with the aforementioned TV regularization methods, the proposed LRTDGS performs the best in removing noise and preserving edges and local details.

Fig. 15 shows the horizontal mean profiles of band 150 for this dataset. From the curves, one can also see that the comparison approaches cannot remove all the noises because some fluctuations occur in their results. LRTDTV and LRTDGS obtain better smoothed curves than those of the other methods. Finally, the proposed LTTDGS achieved better visual results compared to those of LRTDTV. We employed one nonreference image assessment index Q-metric [34] to evaluate the results of real datasets. Table II lists the nonreference index values of the real data test. It can be found that the proposed method obtains the best evaluation results.

C. Discussion

In our LRTDGS model, there are several parameters that need to be discussed, including two regularization parameters λ_1 and λ_2, one penalty parameter β, and Tucker rank (r_1, r_2, r_3) among the three HSI modes. Because the proposed method is similar to LRTDTV [40], the parameters β and Tucker rank (r_1, r_2, r_3) are set as the same as those of LRTDTV for a fair comparison. Thus, we first initialized $\beta = 10^{-2}$ and then updated it as $\beta = \min(\rho\beta, \beta_{\max})$, where $\rho = 1.5$ and $\beta_{\max} = 10^6$. For the Tucker rank (r_1, r_2, r_3), we adopted the same strategy that r_1 and r_2 are set as 80% of spatial size, respectively. In our simulated Indian Pines data, the value of the third-mode r_3 was set to 10 as recommended by LRTDTV and we set r_3 as 5 for the Washington DC Mall dataset. In the real data experiments, the HSI subspace method HySime [55] was utilized to estimate r_3.

To provide a strategy for the determination of regularization parameters λ_1 and λ_2, we provided sensitivity analyses for these two parameters. For the parameter λ_2 of the sparse term, we set it to $\lambda_2 = c/\sqrt{mn}$, where c is a tuning parameter.

We selected cases 1–6 from the simulated Indian Pines data experiment to provide overall sensitivity analyses of λ_1 and λ_2; the average PSNR was used as the evaluation index. Fig. 16 shows the change in PSNR values with respect to these two regularization parameters under different noise cases. The values of λ_2 shown in the figure represent the change of parameter c. From the figures, it can be seen that the change trends of the PSNR values are similar in the different cases, indicating that the regularization parameters are robust for different noisy cases. Moreover, we found that the PSNR values of the LRTDGS method are relatively better when λ_1 changes from 0.1 to 1 and the value of c is within the range of [50, 1000]; thus, we suggest that the parameters λ_1 and c are separately chosen from the aforementioned range according to different noise degrees.

Subsequently, we present a numerical analysis for the convergence of the proposed LRTDGS method. Because of the nonconvexity of Tucker decomposition, our LRTDGS solver is a nonconvex optimization problem. The strategy of updating the penalty parameter β in each iteration has been generally employed in the ALM-based optimization, which can promote convergence of the algorithm [50], [51]. However, it is difficult to theoretically prove the convergence of the algorithm for the nonconvex problem. Thus, the relative change $[(\|\mathcal{X}^{k+1} - \mathcal{X}^k\|_F)/(\|\mathcal{X}^k\|_F)]$ is employed to empirically illustrate the algorithm convergence. Fig. 17 shows the curves of relative change under six different noise cases in the simulated Indian Pines dataset. From the curves, one can clearly see that the relative change value suddenly increases at the beginning, and then converges to zero as the iteration number increases, indicating the convergence of the proposed LRTDGS.

To show the effectiveness of weights for group sparsity constraint, we compare the proposed method with the approach that without the weighted strategy. The proposed method without weighted strategy is denoted as LRTDW. From Table III, we can observe that the proposed method with the weighted strategy achieves better results than that without the weights.

Finally, to further demonstrate the advantage of our proposed method, we selected one state-of-the-art DL method
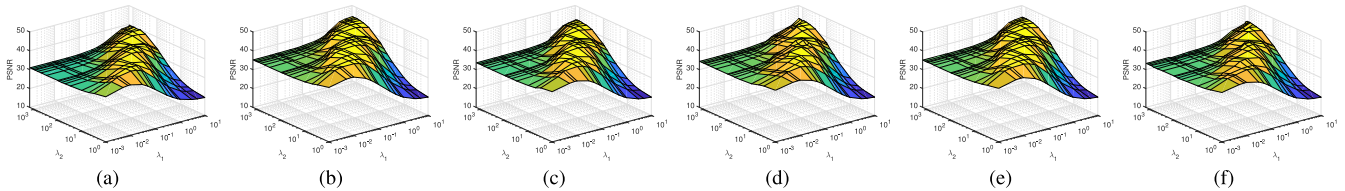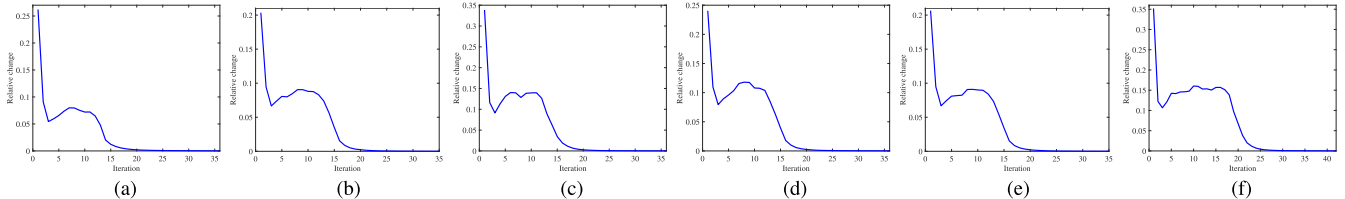
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

CHEN *et al.*: HYPERSPECTRAL IMAGE RESTORATION USING WEIGHTED GROUP SPARSITY-REGULARIZED LOW-RANK TENSOR DECOMPOSITION 13

Fig. 16. Sensitivity analysis between parameters λ_1 and λ_2 using the simulated Indian Pines dataset. (a)–(f) Cases 1–6.

Fig. 17. Relative change values $[(\|\mathcal{X}^{k+1} - \mathcal{X}^k\|_F)/(\|\mathcal{X}^k\|_F)]$ versus the iteration number of the LRTDGS solver in the simulated Indian Pines dataset. (a)–(f) Cases 1–6.

TABLE IV
QUANTITATIVE COMPARISON OF LRTDGS WITH HSID-CNN
IN THE WASHINGTON DC MALL DATASET

Case	Method	PSNR	SSIM	FSIM	ERGAS	SAM
Case 1	HSID-CNN	32.322	0.9134	0.9536	100.056	0.1455
	LRTDGS	**33.961**	**0.9369**	**0.9645**	**82.748**	**0.1269**
Case 2	HSID-CNN	35.313	0.9494	0.9718	76.396	0.1282
	LRTDGS	**37.541**	**0.9727**	**0.9847**	**56.191**	**0.0920**

HSID-CNN [20] for comparison. Since HSID-CNN was mainly designed for Gaussian noise removal, and the model of HSID-CNN was trained on the Washington DC Mall dataset, we adopted cases 1 and 2 simulated experiments on the Washington DC Mall dataset for a fair comparison. Table IV lists the quantitative comparison between LRTDGS and HSID-CNN in the two simulated experiments. It can be observed that our LRTDGS obtains better results than HSID-CNN from quantitative evaluation results.

V. CONCLUSION

In this article, we proposed a group sparsity-regularized low-rank tensor decomposition method for HSI restoration. In this method, group sparsity regularization was utilized to better exploit the shared sparse pattern of the difference image along the spectral dimension and remove Gaussian noise. Moreover, the global spatial–spectral correlation of clean HSI among all bands was described via low-rank tensor decomposition, which can aid in isolating the sparse noise from the clean HSI. We adopted the ALM method to optimize the proposed LRTDGS model, and both simulated and real data experiments validated the efficiency of our proposed LRTDGS method.

In the future, we will attempt to exploit more potential prior for the difference image to further improve the HSI restoration results. In addition, the combination of DL ideas [20], [21], [56] and tensor models is a promising topic for HSI restoration.
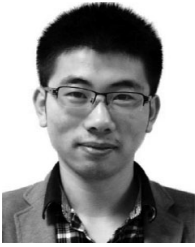
REFERENCES

[1] H. Yuan and Y. Y. Tang, "Spectral–spatial shared linear regression for hyperspectral image classification," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 934–945, Apr. 2017.

[2] N. Yokoya, J. Chanussot, and A. Iwasaki, "Nonlinear unmixing of hyperspectral data using semi-nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1430–1437, Feb. 2014.

[3] Y. Yuan, D. Ma, and Q. Wang, "Hyperspectral anomaly detection by graph pixel selection," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 3123–3134, Dec. 2016.

[4] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4729–4743, Aug. 2014.

[5] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[6] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 1, pp. 65–74, Jan. 1988.

[7] H. Othman and S.-E. Qian, "Noise reduction of hyperspectral imagery using hybrid spatial–spectral derivative-domain wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 2, pp. 397–408, Feb. 2006.

[8] Y.-Q. Zhao and J. Yang, "Hyperspectral image denoising via sparse representation and low-rank constraint," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 296–308, Jan. 2015.

[9] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, p. 11, May 2011.

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

[11] W. He, Q. Yao, C. Li, N. Yokoya, and Q. Zhao, "Non-local meets global: An integrated paradigm for hyperspectral denoising," in *Proc. CVPR*, Jun. 2019, pp. 6868–6877.

[12] Y. Chen, Y. Guo, Y. Wang, D. Wang, C. Peng, and G. He, "Denoising of hyperspectral images using nonconvex low rank matrix approximation," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5366–5380, Sep. 2017.

[13] Y. Xie, Y. Qu, D. Tao, W. Wu, Q. Yuan, and W. Zhang, "Hyperspectral image restoration via iteratively regularized weighted Schatten *p*-norm minimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4642–4659, Aug. 2016.

[14] Y. Xie, S. Gu, Y. Liu, W. Zuo, W. Zhang, and L. Zhang, "Weighted Schatten *p*-norm minimization for image denoising and background subtraction," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4842–4857, Oct. 2016.

[15] Y. Chen, T.-Z. Huang, X.-L. Zhao, and L.-J. Deng, "Hyperspectral image restoration using framelet-regularized low-rank nonnegative matrix factorization," *Appl. Math. Model.*, vol. 63, pp. 128–147, Nov. 2018.

[16] Y. Chen, X. Cao, Q. Zhao, D. Meng, and Z. Xu, "Denoising hyperspectral image with non-IID noise structure," *IEEE Trans. Cybern.*, vol. 48, no. 3, pp. 1054–1066, Mar. 2018.

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

[18] L. Sun, B. Jeon, Y. Zheng, and Z. Wu, "Hyperspectral image restoration using low-rank representation on spectral difference image," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1151–1155, Jul. 2017.

[19] L. Sun, B. Jeon, B. N. Soomro, Y. Zheng, Z. Wu, and L. Xiao, "Fast superpixel based subspace low rank learning method for hyperspectral denoising," *IEEE Access*, vol. 6, pp. 12031–12043, 2018.

[20] Q. Yuan, Q. Zhang, J. Li, H. Shen, and L. Zhang, "Hyperspectral image denoising employing a spatial–spectral deep residual convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1205–1218, Feb. 2019.

[21] Y. Chang, L. Yan, H. Fang, S. Zhong, and W. Liao, "HSI-DeNet: Hyperspectral image restoration via convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 667–682, Feb. 2019.

[22] Q. Zhang, Q. Yuan, J. Li, X. Liu, H. Shen, and L. Zhang, "Hybrid noise removal in hyperspectral imagery with a spatial–spectral gradient network," *IEEE Trans. Geosci. Remote Sens.*, to be published. doi: 10.1109/TGRS.2019.2912909.

[23] A. Karami, M. Yazdi, and A. Z. Asli, "Noise reduction of hyperspectral images using kernel non-negative tucker decomposition," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 487–493, Jun. 2011.

[24] Y. Peng, D. Meng, Z. Xu, C. Gao, Y. Yang, and B. Zhang, "Decomposable nonlocal tensor dictionary learning for multispectral image denoising," in *Proc. CVPR*, Jun. 2014, pp. 2949–2956.

[25] X. Liu, S. Bourennane, and C. Fossati, "Denoising of hyperspectral images using the PARAFAC model and statistical performance analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3717–3724, Oct. 2012.

[26] X. Guo, X. Huang, L. Zhang, and L. Zhang, "Hyperspectral image noise reduction based on rank-1 tensor decomposition," *ISPRS J. Photogram. Remote Sens.*, vol. 83, pp. 50–63, Sep. 2013.

[27] H. Fan, Y. Chen, Y. Guo, H. Zhang, and G. Kuang, "Hyperspectral image restoration using low-rank tensor recovery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 10, pp. 4589–4604, Oct. 2017.

[28] Z. Huang, S. Li, L. Fang, H. Li, and J. A. Benediktsson, "Hyperspectral image denoising with group sparse and low-rank tensor decomposition," *IEEE Access*, vol. 6, pp. 1380–1390, 2018.

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

[30] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D Nonlin. Phenom.*, vol. 60, nos. 1–4, pp. 259–268, Nov. 1992.

[31] Y. Chang, L. Yan, T. Wu, and S. Zhong, "Remote sensing image stripe noise removal: From image decomposition perspective," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7018–7031, Dec. 2016.

[32] Y. Chen, T.-Z. Huang, X.-L. Zhao, L.-J. Deng, and J. Huang, "Stripe noise removal of remote sensing images by total variation regularization and group sparsity constraint," *Remote Sens.*, vol. 9, no. 6, p. 559, Jun. 2017.

[33] J.-H. Yang, X.-L. Zhao, T.-H. Ma, Y. Chen, T.-Z. Huang, and M. Ding, "Remote sensing images destriping using unidirectional hybrid total variation and nonconvex low-rank regularization," *J. Comput. Appl. Math.*, vol. 363, pp. 124–144, Jan. 2020.

[34] Q. Yuan, L. Zhang, and H. Shen, "Hyperspectral image denoising employing a spectral–spatial adaptive total variation model," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3660–3677, Oct. 2012.

[35] H. K. Aggarwal and A. Majumdar, "Hyperspectral image denoising using spatio-spectral total variation," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 442–446, Mar. 2016.

[36] Q. Wang, Z. Wu, J. Jin, T. Wang, and Y. Shen, "Low rank constraint and spatial spectral total variation for hyperspectral image mixed denoising," *Signal Process.*, vol. 142, pp. 11–26, Jan. 2018.

[37] 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.

[38] L. Sun, T. Zhang, Z. Wu, L. Xiao, and B. Jeon, "Hyperspectral mixed denoising via spectral difference-induced total variation and low-rank approximation," *Remote Sens.*, vol. 10, no. 12, p. 1956, Dec. 2018.

[39] 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.

[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] X. Bresson and T. F. Chan, "Fast dual minimization of the vectorial total variation norm and applications to color image processing," *Inverse Problem Imag.*, vol. 2, no. 4, pp. 455–484, 2008.

[42] L. Sun, B. Jeon, Y. Zheng, and Z. Wu, "A novel weighted cross total variation method for hyperspectral image mixed denoising," *IEEE Access*, vol. 5, pp. 27172–27188, 2017.

[43] W. He, N. Yokoya, L. Yuan, and Q. Zhao, "Remote sensing image reconstruction using tensor ring completion and total variation," *IEEE Trans. Geosci. Remote Sens.*, to be published. doi: 10.1109/TGRS.2019.2924017.

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

[45] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 341–354, Jan. 2014.

[46] Y. Chen, T.-Z. Huang, and X.-L. Zhao, "Destriping of multispectral remote sensing image using low-rank tensor decomposition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 12, pp. 4950–4967, Dec. 2018.

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

[48] X. Liu, X. Lu, H. Shen, Q. Yuan, Y. Jiao, and L. Zhang, "Stripe noise separation and removal in remote sensing images by consideration of the global sparsity and local variational properties," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 3049–3060, May 2016.

[49] J. Huang, T.-Z. Huang, L.-J. Deng, and X.-L. Zhao, "Joint-sparse-blocks and low-rank representation for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2419–2438, Apr. 2019.

[50] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, 2010.

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

[52] M. Ishteva, L. De Lathauwer, P.-A. Absil, and S. Van Huffel, "Differential-geometric Newton method for the best rank-(R1, R2, R3) approximation of tensors," *Numer. Algorithms*, vol. 51, no. 2, pp. 179–194, Jun. 2009.

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

[54] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[55] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.

[56] Q. Zhang, Q. Yuan, C. Zeng, X. Li, and Y. Wei, "Missing data reconstruction in remote sensing image with a unified spatial–temporal–spectral deep convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4274–4288, Aug. 2018.

Yong Chen received the B.S. degree from the School of Science, East China University of Technology, Nanchang, China, in 2015. He is currently pursuing the Ph.D. degree with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China.

From 2018 to 2019, he was a Research Intern with the Geoinformatics Unit, RIKEN Center for Advanced Intelligence Project, Tokyo, Japan. His current research interests include remote sensing image processing and sparse optimization.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

CHEN *et al.*: HYPERSPECTRAL IMAGE RESTORATION USING WEIGHTED GROUP SPARSITY-REGULARIZED LOW-RANK TENSOR DECOMPOSITION 15

Wei He (S'14–M'17) received the B.S. degree from the School of Mathematics and Statistics, Wuhan University, Wuhan, China, in 2012, and the Ph.D. degree in surveying, mapping, and remote sensing (LIESMARS) from Wuhan University in 2017.

He is currently a Researcher with the Geoinformatics Unit, RIKEN Center for Advanced Intelligence Project, Tokyo, Japan. His current research interests include image quality improvement, remote sensing image processing, and low-rank representation and deep learning.

Ting-Zhu Huang received the B.S., M.S., and Ph.D. degrees in computational mathematics from the Department of Mathematics, Xi'an Jiaotong University, Xi'an, China.

He is currently a Professor with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China. His current research interests include scientific computation and applications, numerical algorithms for image processing, numerical linear algebra, preconditioning technologies, and matrix analysis with applications.

Prof. Huang is an Editor of *Scientific World Journal*, *Advances in Numerical Analysis*, the *Journal of Applied Mathematics*, the *Journal of Pure and Applied Mathematics: Advances in Applied Mathematics*, and the *Journal of Electronic Science and Technology of China*.

Naoto Yokoya (S'10–M'13) received the M.Eng. and Ph.D. degrees in aerospace engineering from the University of Tokyo, Tokyo, Japan, in 2010 and 2013, respectively.

He is currently a Unit Leader with the RIKEN Center for Advanced Intelligence Project, Tokyo, where he leads Geoinformatics Unit since 2018. He has also been a Visiting Associate Professor with the Tokyo University of Agriculture and Technology, Tokyo, since 2019. He was an Assistant Professor with the University of Tokyo from 2013 to 2017. From 2015 to 2017, he was an Alexander von Humboldt Fellow with the German Aerospace Center (DLR), Oberpfaffenhofen, Germany, and the Technical University of Munich, Munich, Germany. His current research interests include image processing, data fusion, and machine learning for understanding remote sensing images, with applications to disaster management.

Dr. Yokoya was a recipient of the First Place in the 2017 IEEE Geoscience and Remote Sensing Society (GRSS) Data Fusion Contest organized by the Image Analysis and Data Fusion Technical Committee (IADF TC). He has been an Associate Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING since 2018. He has been the Chair of IEEE GRSS IADF TC since 2019 and also the Secretary of IEEE GRSS All Japan Joint Chapter since 2018.