Nonlocal Tensor-Ring Decomposition for Hyperspectral Image Denoising

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

Abstract—Hyperspectral image (HSI) denoising is a fundamental problem in remote sensing and image processing. Recently, nonlocal low-rank tensor approximation-based denoising methods have attracted much attention due to their advantage of being capable of fully exploiting the nonlocal self-similarity and global spectral correlation. Existing nonlocal low-rank tensor approximation methods were mainly based on two common decomposition [Tucker or CANDECOMP/PARAFAC (CP)] methods and achieved the state-of-the-art results, but they are subject to certain issues and do not produce the best approximation for a tensor. For example, the number of parameters for Tucker decomposition increases exponentially according to its dimensions, and CP decomposition cannot better preserve the intrinsic correlation of the HSI. In this article, a novel nonlocal tensor-ring (TR) approximation is proposed for HSI denoising by using TR decomposition to explore the nonlocal self-similarity and global spectral correlation simultaneously. TR decomposition approximates a high-order tensor as a sequence of cyclically contracted third-order tensors, which has strong ability to explore these two intrinsic priors and to improve the HSI denoising results. Moreover, an efficient proximal alternating minimization algorithm is developed to optimize the proposed TR decomposition model efficiently. Extensive experiments on three simulated data sets under several noise levels and two real data sets verify that the proposed TR model provides better HSI denoising results than several state-of-the-art methods in terms of quantitative and visual performance evaluations.

Index Terms—Denoising, hyperspectral image (HSI), nonlocal self-similarity, tensor-ring (TR) decomposition.

I. INTRODUCTION

HYPERSPECTRAL image (HSI) is acquired by imaging spectroscopy over hundreds of bands. This provides

Y. Chen is with the School of Mathematical Sciences/Research Center for Image and Vision Computing, University of Electronic Science and Technology of China, Chengdu 611731, China, and also with the Geoinformatics Unit, RIKEN Center for Advanced Intelligence Project, Tokyo 103-0027, Japan (e-mail: chenyong1872008@163.com).

W. He is with the Geoinformatics Unit, RIKEN Center for Advanced Intelligence Project, Tokyo 103-0027, Japan (e-mail: wei.he@riken.jp).

N. Yokoya is with the Geoinformatics Unit, RIKEN Center for Advanced Intelligence Project, Tokyo 103-0027, Japan, and also with the Department of Electrical and Electronic Engineering, Tokyo University of Agriculture and Technology, Tokyo 183-8538, Japan (e-mail: naoto.yokoya@riken.jp).

T.-Z. Huang and X.-L. Zhao are with the School of Mathematical Sciences/Research Center for Image and Vision Computing, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: tingzhuhuang@126.com; xlzhao122003@163.com).

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

Digital Object Identifier 10.1109/TGRS.2019.2946050

richer spectral information than color images, which is conducive to the fine description of real scenes. However, due to the sensor instability, photon effects, calibration error, and physical mechanism, an HSI is unavoidably contaminated by noise during the imaging process [1]. The noise in the HSI not only destroys the image vision but also influences the succeeding application tasks, such as image fusion [2], [3] and unmixing [4]. Therefore, noise removal from the HSI is a vital topic in the areas of remote sensing and image processing.

Recently, many HSI denoising approaches have been proposed, and they can be approximately grouped into three categories: 2-D image denoising-based methods, 3-D tensor image denoising-based methods, and nonlocal tensor approximation denoising-based methods. The 2-D image denoising method is a classical approach in image processing. The related 2-D approaches, such as total variation (TV) methods [5], [6] and sparse representation [7], can be directly extended to process the HSI separately band by band. These methods can effectively remove low-level noise. However, when the noise is strong, these methods cannot achieve satisfactory denoising results. To remove the different intensities of noise better, the 2-D extended methods were proposed that group similar local patches in the image and achieve better results. These methods include nonlocal sparse representation and low-rank matrix approximation methods [8]–[12]. However, they fail to consider the intrinsic characteristics of global spectral correlation in the HSI, resulting in the spectral distortion of a denoised image. By considering the spectral correlation in the HSI, many researchers have employed low-rank regularization to the HSI by reshaping and stacking all images into a matrix [13]–[19]. Moreover, to avoid computation of singular value decomposition and improve the efficiency, low-rank matrix factorization was also applied to spectral correlation depiction for HSI noise removal [20]–[24]. These methods can effectively preserve the spectral information in the denoising process but fail to preserve the spatial structure.

An HSI is a stack of several 2-D images, and it contains three dimensions: one spectral dimension and two spatial dimensions. Thus, it can be modeled as three-order tensor data. Naturally, tensor-based models can improve the denoising results, because they can fully capture the spatial–spectral correlation of the HSI. There are three kinds of tensor decompositions employed for image denoising: Tucker decomposition [25]–[29], CANDECOMP/PARAFAC (CP) decomposition [1], and tensor singular value decomposition [30]. The advantages of these three tensor decomposition models are that they can simultaneously explore the spatial–spectral correlation between the HSIs within all the bands and better

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 GEOSCIENCE AND REMOTE SENSING
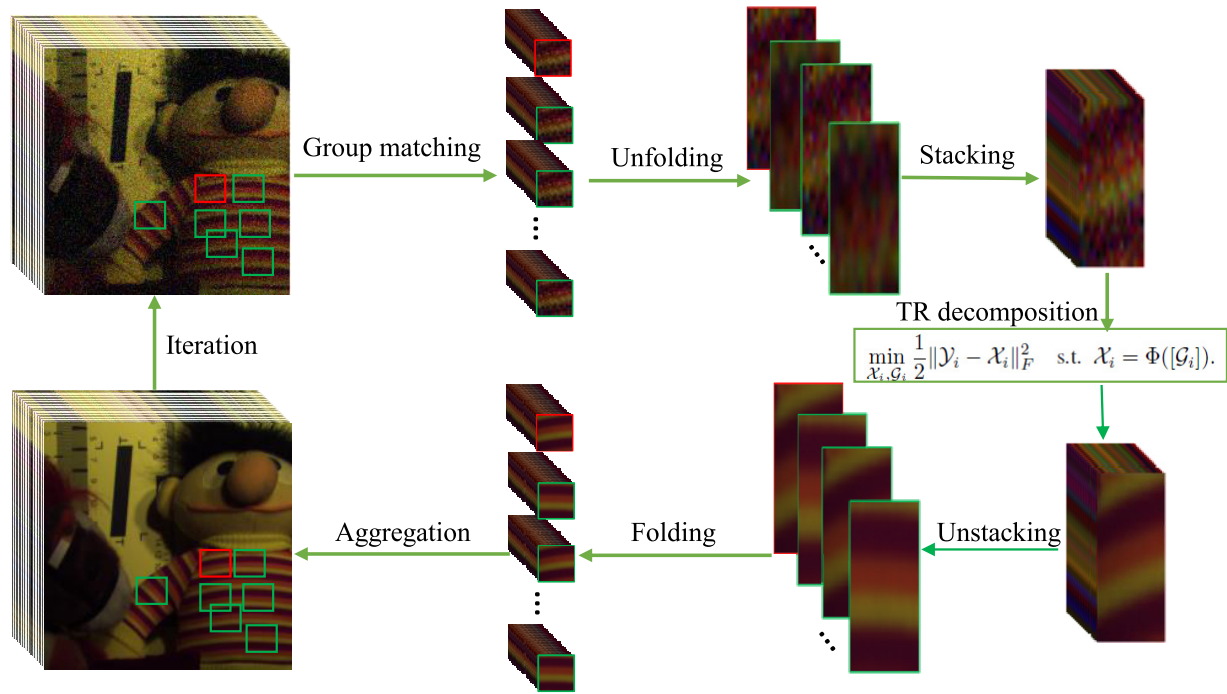

Fig. 1.   Framework of the proposed HSI denoising method.

maintain the spatial–spectral structure in the image. However, the mentioned tensor-based methods ignored the nonlocal self-similarity prior in HSI denoising [31], which is an important prior for improving the denoising performance.

Nonlocal self-similarity indicates the fact that there are many repeated local patterns across the HSI, and the strong correlation exists not only in spectral dimension but also in the nonlocal dimension by grouping the similar patches. The nonlocal similar patch group processing can significantly help the reconstruction of the degraded image [10]. To take full advantage of the nonlocal self-similarity prior in HSI denoising, many nonlocal-based HSI denoising methods have been proposed [31]–[36]. As shown in Fig. 1, nonlocal-based methods try to segment the image into overlapping patches and cluster the similar patches into groups. In [37], the similar nonlocal groups are denoised by a Wiener filter. Peng *et al.* [31] first proposed a tensor dictionary learning (TDL) method to process nonlocal similar groups, which could fully capture the nonlocal self-similarity and global spectral low-rank priors. On the basis of the TDL architecture, many state-of-the-art nonlocal tensor-related approaches were proposed for HSI denoising, including the intrinsic tensor sparsity (ITS) measure method [32], [35] and the hyper-Laplacian regularized unidirectional low-rank tensor recovery problem (LLRT) [33]. With the ITS method, it was thought that the nonlocal group tensors should obtain the property of Tucker and CP decompositions, and a core tensor sparsity-regularized Tucker decomposition model was proposed. For the LLRT method, it was suggested that the correlation of the nonlocal grouped tensor in the spatial and spectral dimensions was not strong enough and only employed the weighted nuclear norm minimization [10] to explore the correlation property along with the nonlocal mode. Although

these two methods achieved good denoising results, the ITS regularization is complex with too many regularizations, while LLRT fails to explore the global spectral information of an HSI. He *et al.* [34] illustrated that the denoising performance can be improved by balancing the correlations between the spectral and nonlocal self-similarities. Thus, there is still much room to improve the performance of HSI denoising.

Nonlocal-based methods are focused on the correlation representation of the grouped three-order tensor, namely local spatial, global spectral, and nonlocal self-similarity correlations. Utilization and balancing of the correlations between the three modes is an important problem. The ITS method is based on the Tucker decomposition and the CP decomposition. Tucker decomposition represents a tensor using one core tensor and a set of factor matrices, and it can capture the correlation of three modes by minimizing the Tucker rank. However, the number of parameters (variables) increases exponentially following its dimensions. Assume that an n-order tensor $\mathcal{X} \in \mathbb{R}^{I \times \cdots \times I}$ and the Tucker rank is $[R, \cdots, R]$; then, the number of variables of the Tucker decomposition is $R^n + nIR$. Moreover, CP decomposition usually achieves satisfactory results with a much larger rank; thus, it cannot better preserve the correlation of the nonlocal mode. Although LLRT is a simple and efficient method, it only employs the low-rank assumption in the nonlocal mode. Therefore, it provides motivation to search for another tensor representation method for balancing the correlations among all the modes to improve the performance of previous nonlocal-related methods. Recently, a new tensor-ring (TR) decomposition was proposed to represent a high-order tensor as a sequence of cyclically contracted third-order tensors [38], [39]. Compared with the Tucker decomposition, the number of

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.*: NLTR DECOMPOSITION FOR HYPERSPECTRAL IMAGE DENOISING

3

variables of TR decomposition is Inr^2 if the TR rank is $[r, \cdots, r]$ (in general, $r < R$). This is much smaller than that of Tucker decomposition. Moreover, because TR factors can be circularly shifted and treated equivalently, it is possible to balance effectively the correlations of all the modes. Based on the efficient representation of a high-order tensor, TR-based methods obtained better results in tensor completion [40]–[43] than those with the Tucker and CP decompositions.

Inspired by the effectiveness and superiority of TR representation, a novel model for HSI denoising through nonlocal TR (NLTR) decomposition was proposed in this article. By introducing the TR decomposition to represent the nonlocal grouped tensor, it was possible to capture and balance the global spectral correlation and nonlocal self-similarity correlation priors of the HSI more efficiently. The contributions of this article include the following.

1) A new tensor representation TR decomposition was introduced to depict the intrinsic spatial, spectral, and nonlocal correlation of similarly grouped tensors simultaneously. The advantage of TR, compared with the Tucker and CP decompositions, was determined by analysis. The new model could be easily extended to handle other high-dimension data-restoration problems.

2) An efficient optimization algorithm proximal alternating minimization (PAM) was employed as the TR factor solver. Extensive experimental results on both the simulated and real data sets demonstrate that the new NLTR method outperforms the state-of-the-art nonlocal methods for HSI denoising in terms of comparison of visual and quantitative indices. The framework of the proposed method is shown in Fig. 1.

The organization of this article is as follows. Section II introduces notations and an HSI denoising framework. The proposed NLTR decomposition and its optimization are formulated in Section III. Section IV reports the simulated and real experimental results and the parameter discussion. Finally, the conclusions for this article are presented in Section V.

II. NOTATIONS AND PROBLEM FORMULATION

A. Notations

In this article, lowercase or uppercase letters (e.g., $i, I \in \mathbb{R}$) are used to denote scalars. Vectors are represented by boldface lowercase letter, e.g., $\mathbf{x} \in \mathbb{R}^I$. Boldface capital letters (e.g., $\mathbf{X} \in \mathbb{R}^{I \times J}$) are employed to represent matrices. A tensor is a multi-dimensional data array, and an n-order tensor ($n \geq 3$) is represented by a calligraphic letter $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_n}$. An element value of $\mathcal{X}$ in position $(i_1, i_2, \cdots, i_n)$ is represented as $\mathcal{X}(i_1, i_2, \cdots, i_n)$ or $x_{i_1 i_2 \cdots i_n}$. Two types of tensor matricization (also called unfolding) expressions are defined in this article. The normal mode-k matricization of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_n}$ is introduced in [44] and represented as $\mathbf{X}_{(k)} \in \mathbb{R}^{I_k \times I_1 \cdots I_{k-1} I_{k+1} \cdots I_n}$. The other mode-$k$ matricization of $\mathcal{X}$, which is introduced in [38] for TR decomposition, is represented as $\mathbf{X}_{[k]} \in \mathbb{R}^{I_k \times I_{k+1} \cdots I_n I_1 \cdots I_{k-1}}$. In contrast, the unfolding matrices $\mathbf{X}_{(k)}$ along the k-mode can be folded as a tensor by $\mathcal{X} = \text{fold}_k(\mathbf{X}_{(k)})$, and the same definition is used for the inverse operation of the second mode-k matricization

$\mathbf{X}_{[k]}$. Moreover, the Frobenius norm of a tensor $\mathcal{X}$ is calculated as $\|\mathcal{X}\|_F = (\sum_{i_1, i_2, \ldots, i_n} (x_{i_1 i_2 \cdots i_n})^2)^{(1/2)}$.

B. HSI Denoising via Tensor Decomposition

The denoising problem is a strong ill-posed problem. Mathematically, HSI denoising can be generally formulated as a regularization model

$$\min_{\mathcal{X}} \frac{1}{2} \|\mathcal{Y} - \mathcal{X}\|_F^2 + \lambda R(\mathcal{X}) \qquad (1)$$

where $\mathcal{Y} \in \mathbb{R}^{M \times N \times B}$ is a noisy HSI with $M \times N$ spatial size and B spectral bands, which is degraded by additive white Gaussian noise; $\mathcal{X} \in \mathbb{R}^{M \times N \times B}$ is a clean HSI to be restored; and λ is a positive regularization parameter used to balance these two terms. The first term is the data fidelity term, and $R(\mathcal{X})$ is the regularization term that characterizes the prior of $\mathcal{X}$.

With model (1), to better and efficiently restore $\mathcal{X}$ from $\mathcal{Y}$, the key problem is transformed to formulate an appropriate regularization term for $\mathcal{X}$. Because the HSI is a 3-D tensor compared with the 2-D image, many tensor-based methods perform HSI denoising by using the tensor decomposition strategy for the HSI data. This includes using Tucker decomposition [25] and CP decomposition [1] to approximate a low-rank tensor. Tucker decomposition is used to find a core tensor and a set of factor matrices, and to formulate them as

$$\mathcal{X} = \mathcal{S} \times_1 \mathbf{U_1} \times_2 \mathbf{U_2} \times_3 \mathbf{U_3} \qquad (2)$$

where $\mathcal{S}$ is the core tensor and $\mathbf{U_k}$ ($k = 1, 2, 3$) are the factor matrices. In addition, CP decomposition is used to decompose a tensor into the sum of rank-1 tensors and can be formulated as

$$\mathcal{X} = \mathbf{A} \circ \mathbf{B} \circ \mathbf{C} = \sum_{k=1}^{R} \mathbf{A}^{(k)} \circ \mathbf{B}^{(k)} \circ \mathbf{C}^{(k)} \qquad (3)$$

where $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are the factor matrices and $\circ$ is the vector out product.

These methods make use of the correlation of different bands in an HSI and can reduce its spatial–spectral redundancy. However, when the tensor representation was directly applied to the whole HSI data, it could not be used to explore the nonlocal self-similarity prior [32]. Recently, several researchers have confirmed that nonlocal methods can achieve the state-of-the-art results for HSI denoising [31]–[33], [35]. Similar to the whole HSI denoising regularization model, the nonlocal HSI noise removal model can be optimized by using the following model:

$$\min_{\mathcal{X}_i} \frac{1}{2} \|\mathcal{Y}_i - \mathcal{X}_i\|_F^2 + \lambda R(\mathcal{X}_i) \qquad (4)$$

where $\mathcal{X}_i \in \mathbb{R}^{P^2 \times B \times K}$ is a third-order tensor stacked by nonlocal similar patches in the ith exemplar patch. P denotes the exemplar patch size and K denotes the number of nonlocal similar patches. After all nonlocal groups $\mathcal{X}_i$ have recovered from $\mathcal{Y}_i$, then $\mathcal{X}_i$ can be aggregated to reconstruct a clean HSI $\mathcal{X}$.

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

4   IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING
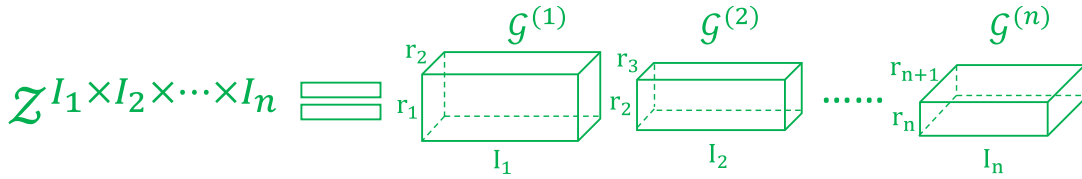

Fig. 2.   Illustration of TR decomposition.

Similar to the model (1), the key problem is the exploring of the prior of $\mathcal{X}_i$ and designing the related regularization to the optimized model (4). By grouping the similar patches together, the global correlation is explored not only in the spectral mode but also in the nonlocal mode. To explore these correlations, TDL [31], ITS [32], [35], and unidirectional low-rank tensor recovery [33] were proposed to describe the prior of $\mathcal{X}_i$ and achieved the state-of-the-art results. TDL is based on low-rank tensor Tucker decomposition and ITS encodes sparsity insights delivered by low-rank tensor Tucker and CP decomposition.

III. PROPOSED NONLOCAL TENSOR-RING DECOMPOSITION FOR HSI DENOISING

Although the mentioned nonlocal tensor HSI denoising methods achieved satisfactory results, there is still room for improvement of these results. Recently, a TR decomposition that represents a high-order tensor as a sequence of cyclically contracted third-order tensors was proposed by Zhao et al. [38]. Compared with Tucker and CP decompositions, TR decomposition can better approximate a high-order tensor, because each tensor factor can be circularly shifted and regarded equivalently under trace operation.

A. Tensor-Ring Decomposition

Compared with other forms of tensor decomposition, the TR form is a more general and effective decomposition. The purpose of TR decomposition is to represent a high-dimensional tensor by circular multilinear products on a series of third-order factor tensors (also called TR factors) [38]. Assume that $\mathcal{Z}$ is an n-dimension tensor with the size of $I_1 \times I_2 \times \cdots \times I_n$. The representation of the TR decomposition of $\mathcal{Z}$ is to search n latent third-order core tensors $\mathcal{G} = \{\mathcal{G}^{(1)}, \mathcal{G}^{(2)}, \cdots, \mathcal{G}^{(n)}\}$, where $\mathcal{G}^{(k)} \in \mathbb{R}^{r_k \times I_k \times r_{k+1}}$. In this case, each elementwise relation of $\mathcal{Z}$ with the core tensor $\mathcal{G}$ can be expressed by

$$\mathcal{Z}(i_1, i_2, \cdots, i_n) = \text{Tr}(\mathbf{G}^{(1)}(i_1)\mathbf{G}^{(2)}(i_2) \cdots \mathbf{G}^{(n)}(i_n))$$
$$= \text{Tr}\left(\prod_{k=1}^{n} \mathbf{G}^{(k)}(i_k)\right) \qquad (5)$$

where $\mathbf{G}^{(k)}(i_k)$ is the i_k lateral slice matrix of $\mathcal{G}^{(k)}$ and $\text{Tr}(\cdot)$ denotes the matrix trace operation. According to the trace operation, the product of these slice matrices should be a square matrix; thus, TR decomposition sets the dimensions of the first and last cores to $r_1 = r_{n+1}$. In this case, vector $r = [r_1, r_2, \cdots, r_n]$ is called the TR-rank. The illustration of TR decomposition is shown in Fig. 2.

Definition 1 (Tensor Multilinear Product [43]): Let $\mathcal{G}^{(k)} \in \mathbb{R}^{r_k \times I_k \times r_{k+1}}$, $k = 1, \cdots, n$ are n third-order tensors in the TR

decomposition, and then, the multilinear product between $\mathcal{G}^{(k)}$ and $\mathcal{G}^{(k+1)}$ is denoted as $\mathcal{G}^{(k,k+1)}$ and calculated as

$$\mathcal{G}^{(k,k+1)}((j_k - 1)I_n + i_k) = \mathcal{G}^{(k)}(i_k)\mathcal{G}^{(k+1)}(j_k) \qquad (6)$$

for $i_k = 1, \cdots, I_k$, $j_k = 1, \cdots, I_{k+1}$ and where $\mathcal{G}^{(k,k+1)} \in \mathbb{R}^{r_k \times I_k I_{k+1} \times r_{k+1}}$.

From Definition 1, it is possible to obtain the multilinear product of all the TR core tensors $[\mathcal{G}] = \prod_{k=1}^{n} \mathcal{G}^{(k)} = \mathcal{G}^{(1,2,\cdots,n)} = \{\mathcal{G}^{(1)}, \mathcal{G}^{(2)}, \cdots, \mathcal{G}^{(n)}\} \in \mathbb{R}^{r_1 \times I_1 I_2 \cdots I_n \times r_1}$. Therefore, the TR decomposition can be rewritten (5) as

$$\mathcal{Z} = \Phi([\mathcal{G}]) \qquad (7)$$

where Φ represents the TR reshaping operator that transforms the third-order tensor $[\mathcal{G}]$ to an n-order tensor $\mathcal{Z}$, i.e., $\Phi : \mathbb{R}^{r_1 \times I_1 I_2 \cdots I_n \times r_1} \rightarrow \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_n}$, and the operator is defined by

$$\mathcal{Z}(i_1, i_2, \cdots, i_n) = \text{Tr}([\mathcal{G}](:, i_1 + (i_2 - 1)I_1 \\ + \cdots + (i_n - 1)I_1 I_2 \cdots I_{n-1}, :)). \qquad (8)$$

Lemma 1 (Circular Dimensional Permutation Invariance [38]): For any n-order tensor $\mathcal{Z} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_n}$, if $\mathcal{Z} = \Phi(\{\mathcal{G}^{(1)}, \mathcal{G}^{(2)}, \cdots, \mathcal{G}^{(n)}\})$ is its TR decomposition, $\overleftarrow{\mathcal{Z}}_k \in \mathbb{R}^{I_k \times I_{k+1} \cdots \times I_n \times I_1 \times \cdots \times I_{k-1}}$ circularly shifts the dimensions of $\mathcal{Z}$ by k. Then, the following relation can be obtained using:

$$\overleftarrow{\mathcal{Z}}_k = \Phi(\{\mathcal{G}^{(k)}, \mathcal{G}^{(k+1)}, \ldots, \mathcal{G}^{(n)}, \mathcal{G}^{(1)}, \ldots, \mathcal{G}^{(k-1)}\}). \qquad (9)$$

With Lemma 1, any TR core tensor can easily be shifted to the first position, indicating that TR decomposition has the ability to capture the correlation of an element in multiple dimensions.

B. Nonlocal Tensor-Ring Decomposition Model

From the definition and properties of TR decomposition, it is clear that it has numerous advantages over other tensor decomposition formats. First, the number of variables in storage is much smaller than that of Tucker decomposition. The number of variables in Tucker decomposition increases exponentially with respect to its original tensor order, indicating that the number of variables to be estimated for TR is obviously decreased. Thus, the estimates with TR will normally be more precise than for Tucker decomposition. Second, TR has the characteristics that each tensor factor can be circularly shifted and regarded equivalently under trace operation, whereas other tensor decomposition techniques fail to preserve this superiority. Third, the decomposition format of TR is tensor to tensor; thus, it has better ability to preserve the original data structure than Tucker or CP decomposition.

Furthermore, we focus on the 2-D spectral matrix analysis and illustrate the advantage of TR decomposition for HSI

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.*: NLTR DECOMPOSITION FOR HYPERSPECTRAL IMAGE DENOISING

5

processing. We extract the spectral signatures of two different materials (denote $S_1 \in R^{B \times k}$ and $S_2 \in R^{B \times k}$) from the HSI data set, where B is the number of band and k is the pixel number of the related materials. Then, we analyze the TR decomposition of the 2-D matrix $S = [S_1, S_2] \in R^{B \times 2k}$. Typically, we adopt $A_1 \in R^{B \times r_1}$ and $A_2 \in R^{B \times r_1}$ to represent the basis of S_1 and S_2, respectively, i.e., $S_1 = A_1 B_1$ and $S_2 = A_2 B_2$. However, matrix/tensor decomposition can learn a global optimal basis of S, ignoring the characteristics of each material. The TR decomposition of S is represented as $S = \Phi\{\mathcal{G}^{(1)}, \mathcal{G}^{(2)}\}$, where $\mathcal{G}^{(1)} \in R^{2 \times B \times r_1}$ is the basis tensor. The core tensor $\mathcal{G}^{(1)}$ has the ability to learn the basis of each material separately. That is to say, TR ring decomposition can precisely obtain the basis of each materials, since TR decomposition can learn several bundles of signatures associated with each basis and robustly estimate basis coefficient. This is similar to the HSI unmixing with spectral variability [45] that relies on the definition of a set of multiple spectral signatures, referred to as endmember bundles, to characterize each endmember class. From the above analysis, we can find that TR decomposition has stronger ability than matrix/tensor decomposition to learn the basis of HSI's different materials.

Based on the superiority of TR decomposition, HSI denoising can be achieved by finding the TR representation of the whole HSI. However, directly using TR decomposition for the whole HSI does not capture the nonlocal self-similarity; thus, TR decomposition was introduced for nonlocal HSI denoising. That is, the regularization term $R(\mathcal{X}_i)$ can be formulated in model (4) as $\|\mathcal{X}_i - \Phi([\mathcal{G}_i])\|_F^2$. By selecting an appropriate parameter, the proposed NLTR decomposition denoising model can be formulated as

$$\min_{\mathcal{X}_i, \mathcal{G}_i} \frac{1}{2} \|\mathcal{Y}_i - \mathcal{X}_i\|_F^2 \quad \text{s.t.} \quad \mathcal{X}_i = \Phi([\mathcal{G}_i]). \tag{10}$$

The proposed NLTR method mainly consists of three steps: grouping nonlocal similar fullband patches, low-rank TR approximation, and aggregating the estimated patches to a clean HSI. First, for each exemplar patch $\mathcal{X}_i$, all similar patches are searched in a local window by the Euclidean distance-based similarity metric. Second, the low-rank tensor approximation is performed by estimating the TR core tensors $\mathcal{G}_i$. Finally, the denoised HSI is recovered by aggregating all the patches. The three steps of the new method are presented in Fig. 1.

C. Optimization Method and Convergence Analysis

The optimization of the NLTR model can be divided into two steps. First, the TR cores $\mathcal{G}_i = \{\mathcal{G}_i^{(1)}, \mathcal{G}_i^{(2)}, \mathcal{G}_i^{(3)}\}$ are estimated by solving the following optimization problem:

$$\min_{\mathcal{G}_i} \frac{1}{2} \|\mathcal{Y}_i - \Phi([\mathcal{G}_i])\|_F^2. \tag{11}$$

Once the TR cores $\mathcal{G}_i$ are obtained, $\mathcal{X}_i$ is computed as $\mathcal{X}_i = \Phi([\mathcal{G}_i])$. The key problem is how to optimize the model (11). Clearly, the problem of (11) is not jointly convex for $\mathcal{G}_i$, but it is convex for $\mathcal{G}_i = \{\mathcal{G}_i^{(1)}, \mathcal{G}_i^{(2)}, \mathcal{G}_i^{(3)}\}$ independently. To optimize the nonconvex problem effectively, PAM is applied [46]–[49].

Before optimizing (11), the problem of (11) is transformed into its equivalent form by introducing the proximal operator [50], and the problem can be rewritten as

$$\min_{\mathcal{G}_i} \frac{1}{2} \|\mathcal{Y}_i - \Phi([\mathcal{G}_i])\|_F^2 + \frac{\rho}{2} \|\mathcal{G}_i - (\mathcal{G}_i)^j\|_F^2 \tag{12}$$

where ρ denotes a positive proximal parameter and $(\mathcal{G}_i)^j$ is the result of the last iteration of $\mathcal{G}_i$. Then, it becomes possible to update alternately each core $\mathcal{G}_i^{(k)}(k = 1, 2, 3)$ as follows:

$$\begin{cases} (\mathcal{G}_i^{(1)})^{j+1} = \min_{\mathcal{G}_i^{(1)}} \frac{1}{2}\|\mathcal{Y}_i - \Phi(\{\mathcal{G}_i^{(1)}, (\mathcal{G}_i^{(2)})^j, (\mathcal{G}_i^{(3)})^j\})\|_F^2 \\ \qquad + \frac{\rho}{2}\|\mathcal{G}_i^{(1)} - (\mathcal{G}_i^{(1)})^j\|_F^2 \\ (\mathcal{G}_i^{(2)})^{j+1} = \min_{\mathcal{G}_i^{(2)}} \frac{1}{2}\|\mathcal{Y}_i - \Phi(\{(\mathcal{G}_i^{(1)})^{j+1}, \mathcal{G}_i^{(2)}, (\mathcal{G}_i^{(3)})^j\})\|_F^2 \\ \qquad + \frac{\rho}{2}\|\mathcal{G}_i^{(2)} - (\mathcal{G}_i^{(2)})^j\|_F^2 \\ (\mathcal{G}_i^{(3)})^{j+1} = \min_{\mathcal{G}_i^{(3)}} \frac{1}{2}\|\mathcal{Y}_i - \Phi(\{(\mathcal{G}_i^{(1)})^{j+1}, (\mathcal{G}_i^{(2)})^{j+1}, \mathcal{G}_i^{(3)}\})\|_F^2 \\ \qquad + \frac{\rho}{2}\|\mathcal{G}_i^{(3)} - (\mathcal{G}_i^{(3)})^j\|_F^2. \end{cases} \tag{13}$$

According to Lemma 1, each core tensor is shifted to the first position when optimizing this core tensor, which gives

$$\mathbf{Z}_{<k>} = (\overleftarrow{\mathbf{Z}}_k)_{(1)} = \mathbf{G}_{(2)}^{(k)} \left(\mathbf{G}_{<2>}^{(k+1, \cdots, 1, \dots, k-1)}\right)^T. \tag{14}$$

Based on (14), the least square problem (13) can be solved as follows:

$$\begin{cases} (\mathcal{G}_i^{(1)})^{j+1} = \text{fold}_2\left((\mathbf{Y}_{i,<1>}(\mathbf{G}_{i,<2>}^{(2,3)})^j + \rho(\mathbf{G}_{i,(2)}^{(1)})^j\right) \\ \qquad \times \left((\mathbf{G}_{i,<2>}^{(2,3)})^{j,T}(\mathbf{G}_{i,<2>}^{(2,3)})^j + \rho\mathbf{I}\right)^{-1}\right) \\ (\mathcal{G}_i^{(2)})^{j+1} = \text{fold}_2\left((\mathbf{Y}_{i,<2>}(\mathbf{G}_{i,<2>}^{(3,1)})^j + \rho(\mathbf{G}_{i,(2)}^{(2)})^j\right) \\ \qquad \times \left((\mathbf{G}_{i,<2>}^{(3,1)})^{j,T}(\mathbf{G}_{i,<2>}^{(3,1)})^j + \rho\mathbf{I}\right)^{-1}\right) \\ (\mathcal{G}_i^{(3)})^{j+1} = \text{fold}_2\left((\mathbf{Y}_{i,<3>}(\mathbf{G}_{i,<2>}^{(1,2)})^j + \rho(\mathbf{G}_{i,(2)}^{(3)})^j\right) \\ \qquad \times \left((\mathbf{G}_{i,<2>}^{(1,2)})^{j,T}(\mathbf{G}_{i,<2>}^{(1,2)})^j + \rho\mathbf{I}\right)^{-1}\right) \end{cases} \tag{15}$$

where $\mathbf{I}$ is the identity matrix. After the optimization of core tensors $\mathcal{G}_i$, $\mathcal{X}_i$ can be updated as follows:

$$\mathcal{X}_i = \Phi([\mathcal{G}_i]). \tag{16}$$

The whole HSI denoising procedure is summarized in Algorithm 1.

According to the solver framework in [46], the proposed algorithm can converge to a critical point of the objection function. The theoretical proof of convergence is described in the Appendix.

D. Computational Complexity

Considering Algorithm 1 for an input-noisy HSI $\mathcal{Y} \in \mathbb{R}^{M \times N \times B}$, the total number of exemplar groups is $S = O(MN)$. The main computation complexity of each exemplar group is to update $\mathcal{G}_i$ in Step 8. Assuming that the size of each group is $\mathcal{Y}_i \in \mathbb{R}^{I \times I \times I}$ and the TR rank is set as $r_1 = r_2 = r_3 = r$, the cost of each exemplar group updating in each iteration is $O(r^6 + I^2 r^4 + I^3 r^2)$. For quite large spatial size, the computation cost is extremely high. However, the denoising can be performed on the S groups using parallel

Fig. 3.   Five simulated HSIs selected from the CAVE data set. The color image is composed of bands 31, 22, and 1. (a) Balloons. (b) Toy. (c) Cloth. (d) Feather. (e) Food.

Algorithm 1 NLTR Method for HSI Denoising

Require: Degraded HSI $\mathcal{Y}$, TR rank $r = [r_1, r_2, r_3]^T$, parameters ρ and δ.
1: Initialize: Let $\mathcal{X}^{(0)} = \mathcal{Y}$, $\mathcal{Y}^{(0)} = \mathcal{Y}$.
2: **for** $t = 1 : T$ **do**
3:    Calculate $\mathcal{Y}^{(t)} = \mathcal{Y}^{(t-1)} + \delta(\mathcal{Y} - \mathcal{X}^{(t-1)})$.
4:    **for** each exemplar patch $\mathbf{Y}_i$ **do**
5:       Grouping non-local similar patches to form tensor $\mathcal{Y}_i$.
6:       Randomly initialize for core tensors $\mathcal{G}_i$.
7:       **while** not converged **do**
8:          Alternately update $\mathcal{G}_i = \{\mathcal{G}_i^{(1)}, \mathcal{G}_i^{(2)}, \mathcal{G}_i^{(3)}\}$ via (15).
9:       **end while**
10:      Compute $\mathcal{X}_i = \Phi([\mathcal{G}_i])$.
11:   **end for**
12:   Aggregate all $\mathcal{X}_i$ to achieve the denoised HSI $\mathcal{X}^{(t)}$.
13: **end for**
Ensure: Denoised HSI result $\mathcal{X}^{(T)}$.

---

computing. Moreover, the group processing of each exemplar patch is then searched on a 2-D matrix that is obtained by averaging each band of $\mathcal{Y}$ [33]. Therefore, the computational cost can be significantly reduced.

IV. EXPERIMENTAL RESULTS

In the experimental tests, simulated and real experiments were performed to illustrate the effectiveness of the NLTR method for HSI denoising. The following nine state-of-the-art and representative denoising methods were employed for comparison: bandwise BM3D [51],[1] LRMR [52], BM4D [37], LRTA [25], PARAFAC [1], LRTDTV [28], TDL [31],[2] KBR (an extension of ITS) [35], and LLRT [33].[3] The codes of the comparison methods are available from the authors' homepages. Moreover, all parameter selections for the comparison approaches were determined from the authors' codes or from the suggestion in their articles to obtain the best performance. To help reproduce all of the experimental results for HSI denoising, the NLTR code will be published on one of the author's homepage.[4] The patch size $P = 6$, the number of

similar patches $K = 200$, and proximal parameter $\rho = 1$ were used in all the experiments. The detailed discussions of parameter selection for NLTR are presented in Section IV-D.

A. Simulated Experiments

In the simulated experiments, three popular data sets were selected to simulate a noisy HSI. The first was the Columbia MSI data set, (CAVE[5]) which includes 32 real-world objects and materials. Each data set contained spatial size 512×512 and spectral dimension 31. In the test, five MSIs were selected from which to extract subimages with a size of $200 \times 200 \times 31$ to perform the experiments shown in Fig. 3. The second and third HSIs were selected from the hyperspectral Pavia City Center data set (PaC[6]) and the Washington DC Mall data set, (WDC[7]) respectively. Two subimages with size $200 \times 200 \times 50$ were extracted from these two hyperspectral data sets as the clean image. To simulate different noisy levels, zero-mean additive Gaussian noise was simulated with noise variance σ of 10, 30, 50, or 100 for all the clean data sets.

To evaluate the denoising results of all methods, five objective quantitative indices were computed for comparison, including the peak signal-to-noise ratio (PSNR), the structure similarity (SSIM) [53], the feature similarity (FSIM) [54], the erreur relative globale adimensionnelle de synthese (ERGAS) [55], and the spectral angle mapper (SAM). PSNR and SSIM are two common indices for image restoration. FSIM illustrates the perceptual consistency compared with the ground truth. ERGAS and SAM are based on the spectral evaluation index. In general, larger PSNR, SSIM, and FSIM indicate better denoising results. In contrast, better results are needed to obtain smaller ERGAS and SAM. The image denoising results are shown for visual comparison.

1) Experimental Results for CAVE Data Set: Figs. 4 and 5 show the visual results for the CAVE Toy and Feather MSIs at two different noise intensities, respectively. LRMR failed to remove the noise in both the cases, because it used matrix low-rank to perform the denoising processing. With the noise variance $\sigma = 50$, other methods were able to remove the noise, but LRTA, PARAFAC, and LRTDTV produced

---

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.*: NLTR DECOMPOSITION FOR HYPERSPECTRAL IMAGE DENOISING   7
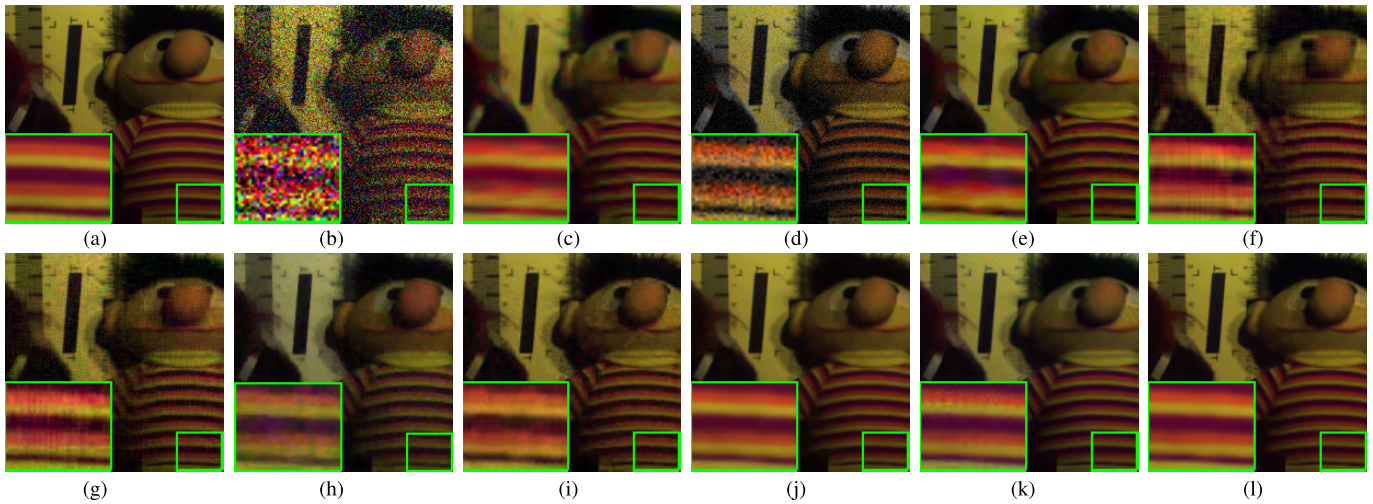

Fig. 4.   Denoised results of CAVE-Toy image of the noise variance $\sigma = 50$. (a) Original. (b) Noisy. (c) BM3D. (d) LRMR. (e) BM4D. (f) LRTA. (g) PARAFAC. (h) LRTDTV. (i) TDL. (j) KBR. (k) LLRT. (l) NLTR.
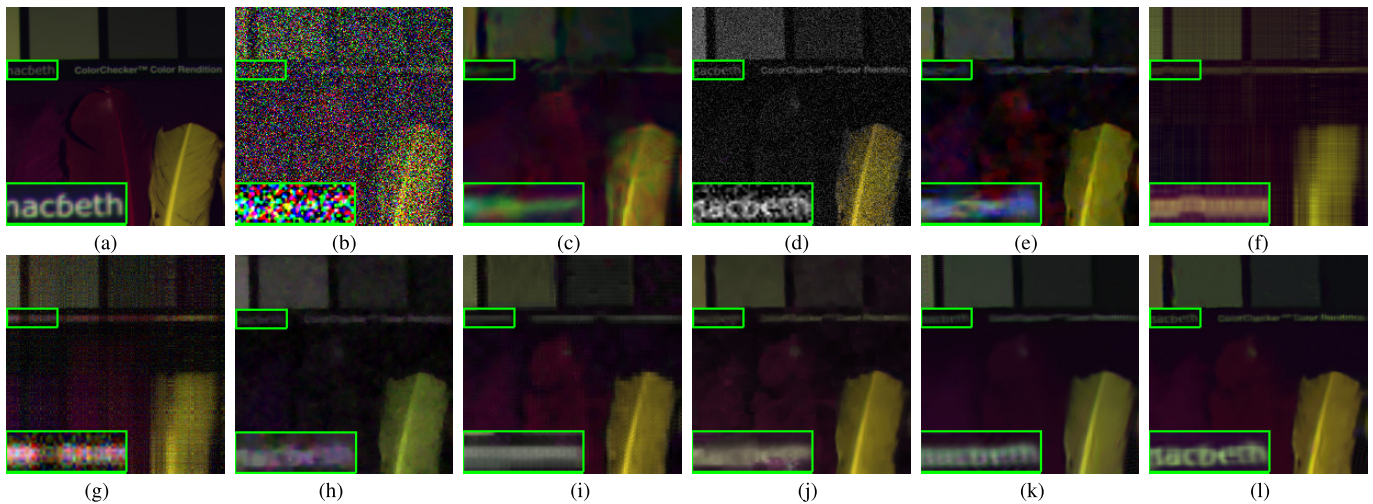

Fig. 5.   Denoised results of the CAVE-Feather image on the noise variance $\sigma = 100$. (a) Original. (b) Noisy. (c) BM3D. (d) LRMR. (e) BM4D. (f) LRTA. (g) PARAFAC. (h) LRTDTV. (i) TDL. (j) KBR. (k) LLRT. (l) NLTR.

some artifacts. Moreover, BM3D and BM4D destroyed the image details. These visual conditions can be observed in the enlarged box. Fig. 5 shows the results of high noise variance. BM3D, LRMR, BM4D, LRTA, PARAFAC, and LRTDTV could not completely remove the noise completely, and the image textures were not preserved. The nonlocal-based methods TDL, KBP, and LLRT effectively eliminated noise, but image detail was lost. From the enlarged box, the new nonlocal-based TR decomposition method NLTR was found to achieve the best visual denoising results, thereby removing the noise effectively and preserving most of the image details and textures.

Table I presents the quantitative results from comparison of the denoising approaches with the CAVE data set at four different noise levels, calculated by averaging five MSI scenes for each noise level. We highlight the top three algorithms by using three different colors (red, green, and blue). Among the three colors, the best performance for each quality index is highlighted in red, the second-best performance is highlighted

in green, and the third-best performance is highlighted in blue. From Table I, nonlocal-based tensor decomposition methods show better results than the other methods. Moreover, the NLTR approach achieves better improvement of the denoising results than the state-of-the-art nonlocal-based tensor methods TDL, KBP, and LLRT, which indicates the superiority of TR decomposition for tensor representation. Fig. 6 shows the PSNR values of each band on the CAVE data set at four different noise levels, which are also computed by averaging the five MSIs. The results again demonstrate that nonlocal-based methods obtained superior results, and that the NLTR approach achieved the best results in most of the bands.

B. Experimental Results for the PaC and WDC Data Sets

Figs. 7 and 8 show the denoising results for the PaC and WDC data sets, respectively. BM3D removed the noise band by band, but the image details were destroyed.

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 GEOSCIENCE AND REMOTE SENSING

TABLE I

QUANTITATIVE RESULTS OF DIFFERENT COMPARISON METHODS ON THE CAVE DATA SET UNDER DIFFERENT NOISE VARIANCES

| σ	Index	Noisy	BM3D	LRMR	BM4D	LRTA	PARAFAC	LRTDTV	TDL	KBR	LLRT	NLTR
	PSNR	28.132	39.415	38.725	42.789	39.917	32.653	39.498	42.597	43.803	45.074	45.178
	SSIM	0.539	0.963	0.928	0.981	0.952	0.846	0.969	0.979	0.984	0.988	0.988
10	FSIM	0.740	0.968	0.955	0.982	0.961	0.906	0.976	0.984	0.986	0.988	0.989
	ERGAS	197.255	53.664	58.526	35.428	49.406	116.052	53.695	36.618	32.481	27.481	27.477
	SAM	0.386	0.081	0.119	0.059	0.089	0.161	0.089	0.057	0.048	0.040	0.039
	PSNR	18.589	33.379	30.643	36.459	33.785	32.163	34.163	36.399	38.359	38.854	39.309
	SSIM	0.171	0.880	0.687	0.932	0.857	0.795	0.905	0.935	0.955	0.962	0.964
30	FSIM	0.441	0.912	0.831	0.949	0.904	0.876	0.937	0.952	0.965	0.968	0.970
	ERGAS	591.764	105.908	147.108	72.946	98.315	121.613	101.266	73.738	59.503	56.046	53.671
	SAM	0.748	0.153	0.263	0.124	0.148	0.194	0.162	0.099	0.079	0.072	0.069
	PSNR	14.152	30.980	26.690	33.560	31.110	28.864	31.645	33.478	35.608	35.559	36.728
	SSIM	0.080	0.817	0.497	0.877	0.780	0.620	0.841	0.896	0.929	0.927	0.942
50	FSIM	0.322	0.873	0.723	0.920	0.865	0.786	0.901	0.926	0.947	0.946	0.957
	ERGAS	986.273	139.467	231.920	102.019	134.111	174.271	131.471	103.699	80.879	81.344	71.202
	SAM	0.942	0.192	0.380	0.176	0.181	0.301	0.211	0.121	0.104	0.107	0.090
	PSNR	8.132	27.638	21.275	29.742	27.454	23.794	28.264	30.199	30.995	32.305	32.898
	SSIM	0.023	0.708	0.262	0.747	0.643	0.329	0.724	0.801	0.841	0.874	0.884
100	FSIM	0.199	0.816	0.551	0.860	0.796	0.612	0.829	0.868	0.894	0.910	0.925
	ERGAS	1972.546	201.051	433.415	157.884	204.139	317.231	191.158	150.094	136.462	118.671	109.310
	SAM	1.184	0.269	0.608	0.270	0.244	0.493	0.297	0.168	0.175	0.142	0.132
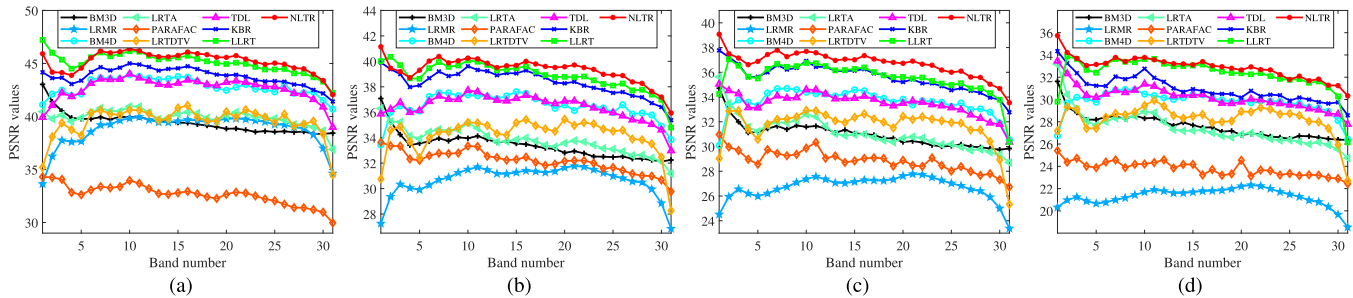
Fig. 6.   PSNR values of each band of the CAVE data set at different noise levels. (a) $\sigma = 10$. (b) $\sigma = 30$. (c) $\sigma = 50$. (d) $\sigma = 100$.

The noise was eliminated by BM4D and KBR, but they obtained blurry results. Although LRTA, PARAFAC, LRT-DTV, and TDL obtained acceptable results at low noise levels, they could not remove noise at high noise levels. In comparison, LLRT and the proposed NLTR method achieved superior results competitive with the other methods.

Tables II and III show the comparison of quantitative results for the PaC and WDC data sets, respectively. From Tables II and III, the NLTR method shows a slight improvement over the comparison methods at low noise levels. However, when the HSIs are seriously degraded by high noise levels, the proposed method demonstrates significant improvement in these four quantitative indices. Moreover, it can again be confirmed that the nonlocal-based methods outperformed the other methods, because they could explore more correlation information in the image. Fig. 9 shows the PSNR value of each band for the PaC and WDC data sets. It is evident that NLTR achieves much higher PSNR for almost every band except $\sigma = 30$ in the WDC data set, which indicates the robustness of the proposed method. In summary, extensive simulation experiments demonstrated that the NLTR method outperforms existing denoising approaches.

C. Real Experiments

In the above, a series of simulated experiments was performed to illustrate the effectiveness of the NLTR method; however, in real scenes, noise degradation is very complex. To demonstrate the effectiveness of the NLTR method in handling actual noisy HSIs, two remarkable noisy data sets were selected for testing. The first data set was collected by the NASA AVIRIS instrument over the Indian Pines test site,[8] and the original size was $145 \times 145 \times 220$. Some bands of this data set are seriously degraded with complicated noise. The second data set was collected by the reflective optics system imaging spectrometer (ROSIS) of the Pavia University,[9] and the data contained 610×340 spatial pixels with 103 spectral bands. Because the first few bands are seriously degraded by noise, several bands with spatial size 340×340 were extracted to test.

Figs. 10 and 11 show the denoised results on the Indian Pines and Pavia University data sets, respectively. As presented

[8] https://engineering.purdue.edu/ biehl/MultiSpec/hyperspectral.html
[9] http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes

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

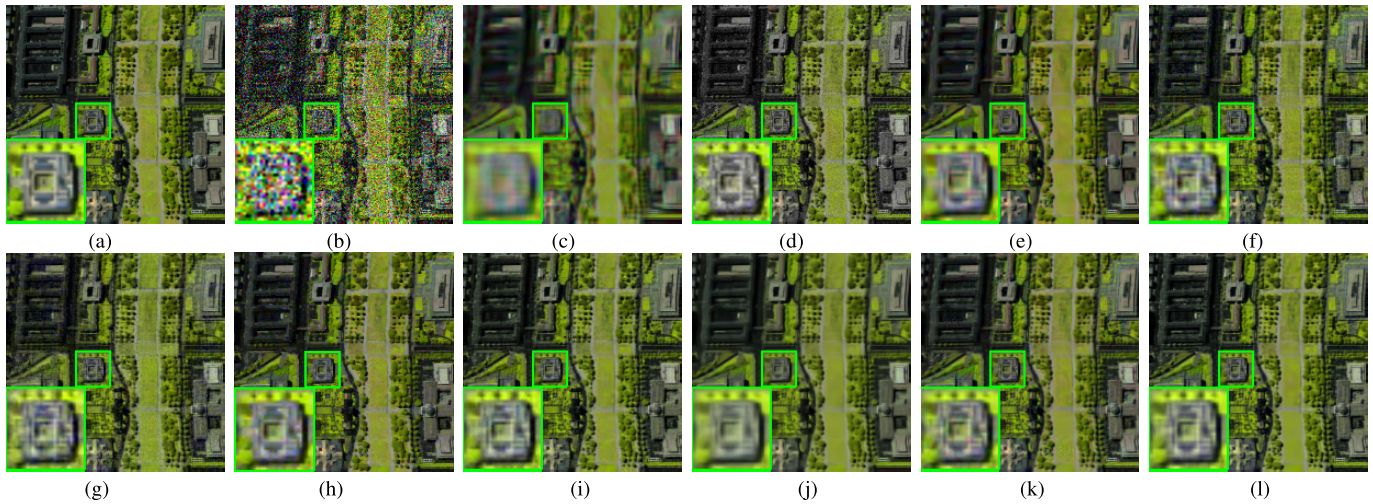CHEN *et al.*: NLTR DECOMPOSITION FOR HYPERSPECTRAL IMAGE DENOISING 9

Fig. 7. Denoised results of WDC image of the noise variance $\sigma = 50$. (a) Original. (b) Noisy. (c) BM3D. (d) LRMR. (e) BM4D. (f) LRTA. (g) PARAFAC. (h) LRTDTV. (i) TDL. (j) KBR. (k) LLRT. (l) NLTR.

TABLE II

COMPARISON OF QUANTITATIVE RESULTS OF DIFFERENT METHODS USED ON THE PaC DATA SET UNDER DIFFERENT NOISE VARIANCES

σ	Index	Noisy	BM3D	LRMR	BM4D	LRTA	PARAFAC	LRTDTV	TDL	KBR	LLRT	NLTR
	PaC dataset											
10	PSNR	28.136	32.516	40.221	38.923	37.712	28.596	39.462	40.913	39.035	41.478	41.856
	SSIM	0.835	0.934	0.988	0.984	0.977	0.854	0.986	0.990	0.985	0.991	0.992
	FSIM	0.910	0.955	0.992	0.989	0.983	0.890	0.991	0.993	0.990	0.994	0.994
	ERGAS	133.080	80.126	33.865	38.396	44.775	126.595	36.326	30.837	38.003	28.866	27.685
	SAM	0.264	0.130	0.057	0.072	0.086	0.129	0.059	0.049	0.047	0.043	0.042
30	PSNR	18.594	26.913	32.686	32.279	31.462	28.898	33.348	34.048	33.336	34.100	34.836
	SSIM	0.418	0.785	0.938	0.928	0.915	0.867	0.943	0.954	0.944	0.956	0.963
	FSIM	0.711	0.864	0.963	0.951	0.935	0.902	0.962	0.971	0.964	0.971	0.975
	ERGAS	399.241	152.675	81.643	82.442	90.936	122.927	73.737	67.441	73.054	67.023	61.934
	SAM	0.571	0.175	0.115	0.120	0.126	0.135	0.108	0.072	0.076	0.075	0.069
50	PSNR	14.157	24.509	28.609	29.274	28.181	28.053	30.292	30.647	30.470	30.398	32.507
	SSIM	0.220	0.660	0.878	0.863	0.835	0.839	0.889	0.907	0.894	0.898	0.938
	FSIM	0.583	0.793	0.929	0.909	0.882	0.888	0.929	0.940	0.931	0.935	0.958
	ERGAS	665.401	201.320	130.850	116.491	132.370	134.918	106.564	99.795	101.515	102.471	80.617
	SAM	0.774	0.200	0.173	0.152	0.141	0.159	0.148	0.082	0.080	0.116	0.086
100	PSNR	8.136	21.663	23.926	25.478	24.454	23.992	26.486	26.899	27.166	27.825	29.218
	SSIM	0.066	0.452	0.702	0.709	0.664	0.663	0.769	0.786	0.793	0.828	0.874
	FSIM	0.406	0.685	0.850	0.817	0.803	0.812	0.864	0.873	0.864	0.891	0.917
	ERGAS	1330.802	279.357	218.440	180.259	202.749	214.266	174.358	153.152	148.415	137.849	117.408
	SAM	1.064	0.234	0.243	0.196	0.155	0.268	0.222	0.094	0.087	0.116	0.113

in the denoised results of Fig. 10, it can be seen that the BM3D method blurred the image to remove the noise. With LRMR, LRTA, and TDL, the results included persistent noise in the images. PARAFAC and LRTDTV removed the noise but introduced some artifacts. Compared with the other methods, KBR and LLTR obtained better results, but some details were not well preserved. In contrast, NLTR simultaneously removed the noise and preserved the image details. The Pavia University data set was degraded by lower noise levels than with the Indian Pines data set, but BM3D, BM4D, LRTA, and TDL could not completely remove the noise. TDL achieved satisfactory denoising results in the simulation experiments, but failed to achieve such results using real data. LRMR effectively removed the noise using noise-free band information, but the image contrast was changed. Although LRTDTV and LLRT obtained better results than the above, the image details were blurred (see the enlarged box). After denoising, KBR and NLTR provided the best denoising results compared with all other methods. In summary, even without the ground-truth HSI as a reference, visual comparison also demonstrated the superiority of the proposed NLTR method for denoising real data. Moreover, we employed one non-reference image-assessment index Q-metric [56] to evaluate the results of real data sets. Table IV lists the nonreference index values of the real data sets. Combining the visual comparison and the nonreference image-assessment index, we can observe that the proposed method obtains the best results.

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

10   IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING

TABLE III

COMPARISON OF QUANTITATIVE RESULTS OF DIFFERENT METHODS USED ON THE WDC DATA SET UNDER DIFFERENT NOISE VARIANCES

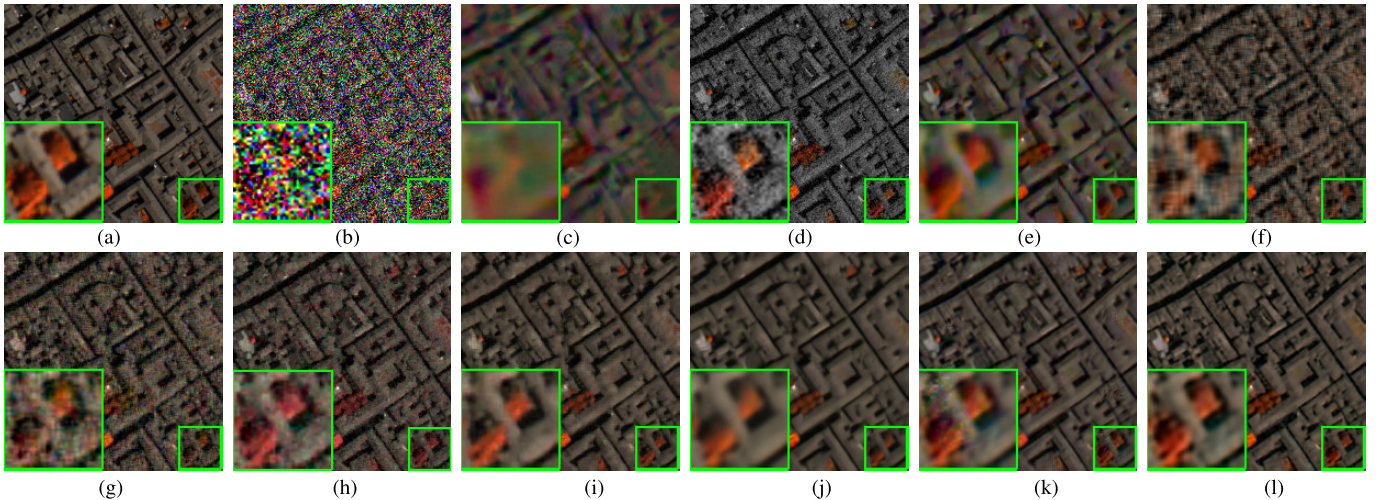| σ	Index	Noisy	BM3D	LRMR	BM4D	LRTA	PARAFAC	LRTDTV	TDL	KBR	LLRT	NLTR
						WDC dataset						
10	PSNR	28.136	31.180	38.427	37.050	36.358	28.020	37.603	39.623	36.876	40.016	40.513
	SSIM	0.874	0.939	0.986	0.982	0.975	0.881	0.983	0.989	0.982	0.990	0.991
	FSIM	0.935	0.960	0.991	0.988	0.984	0.927	0.990	0.993	0.989	0.993	0.994
	ERGAS	115.751	80.986	37.121	41.801	48.162	123.340	40.592	33.240	43.124	31.778	29.744
	SAM	0.216	0.118	0.058	0.070	0.090	0.116	0.063	0.048	0.050	0.043	0.040
30	PSNR	18.594	25.154	31.461	30.257	30.273	27.588	31.964	32.402	31.104	32.624	32.143
	SSIM	0.519	0.783	0.937	0.923	0.917	0.870	0.944	0.950	0.937	0.952	0.949
	FSIM	0.777	0.876	0.966	0.953	0.949	0.922	0.965	0.969	0.963	0.970	0.970
	ERGAS	347.252	161.789	84.866	90.899	94.642	130.590	78.404	74.956	83.217	72.100	78.177
	SAM	0.485	0.162	0.112	0.119	0.124	0.128	0.100	0.080	0.064	0.075	0.074
50	PSNR	14.157	22.740	28.117	27.276	27.263	26.832	29.086	29.016	28.237	28.654	30.279
	SSIM	0.310	0.643	0.883	0.857	0.856	0.849	0.895	0.903	0.881	0.894	0.925
	FSIM	0.664	0.801	0.940	0.916	0.915	0.912	0.939	0.944	0.930	0.939	0.955
	ERGAS	578.753	213.566	126.055	127.677	132.247	142.074	111.653	109.003	116.116	110.893	95.086
	SAM	0.668	0.174	0.144	0.144	0.137	0.142	0.130	0.089	0.068	0.112	0.086
100	PSNR	8.136	20.240	23.149	23.647	23.552	23.597	25.471	24.926	24.516	25.841	27.306
	SSIM	0.110	0.451	0.739	0.699	0.721	0.731	0.793	0.779	0.744	0.816	0.865
	FSIM	0.489	0.696	0.877	0.833	0.853	0.860	0.889	0.880	0.847	0.897	0.920
	ERGAS	1157.506	284.896	225.839	193.296	198.987	201.019	193.619	175.089	178.316	153.339	130.938
	SAM	0.950	0.213	0.216	0.179	0.145	0.222	0.200	0.103	0.090	0.111	0.102

Fig. 8.   Denoised results of a PaC image of the noise variance $\sigma = 100$. (a) Original. (b) Noisy. (c) BM3D. (d) LRMR. (e) BM4D. (f) LRTA. (g) PARAFAC. (h) LRTDTV. (i) TDL. (j) KBR. (k) LLRT. (l) NLTR.

TABLE IV

BLIND ASSESSMENT INDEX Q-METRIC COMPARISON ON THE REAL DATA

Dataset	BM3D	LRMR	BM4D	LRTA	PARAFAC	LRTDTV	TDL	KBR	LLRT	NLTR
Indian	0.0397	0.0367	0.0396	0.0377	0.0361	0.0377	0.0376	0.0398	0.0390	0.0398
Pavia	0.0208	0.0198	0.0210	0.0208	0.0135	0.0192	0.0209	0.0211	0.0202	0.0213

D. Discussion

In the NLTR model, there are several parameters that need to be determined before performing Algorithm 1, including patch size P, similar patch numbers K, proximal parameter ρ, iterative regularization parameter δ, and TR rank $r = [r_1, r_2, r_3]$. For fair comparison, the iterative regularization parameter was set to $\delta = 0.1$ as introduced in [32] and [35] for all the experiments. Although there are many other parameters

in the model, all are very robust and could be fixed in all experiments. Next, the sensibility of these parameters will be analyzed. A subimage of Toy from the CAVE data set and the PaC data set was selected as the experimental object, and PSNR was employed as the evaluation measure to analyze the influence of the parameters.

1) Analysis of Patch Size P: Fig. 12(a) shows the PSNR values with different patch sizes and noise levels.

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.*: NLTR DECOMPOSITION FOR HYPERSPECTRAL IMAGE DENOISING
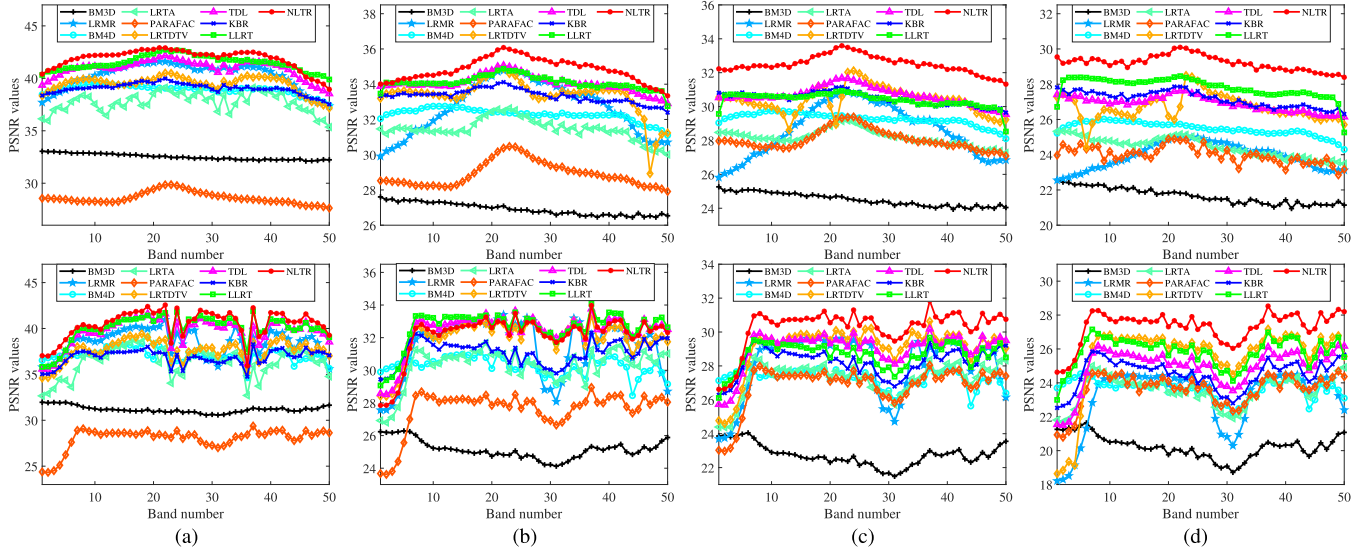
11

Fig. 9. PSNR values of each band in the PaC and WDC data sets at different noise levels. Top: PaC data set, Bottom: WDC data set. (a) $\sigma = 10$. (b) $\sigma = 30$. (c) $\sigma = 50$. (d) $\sigma = 100$.
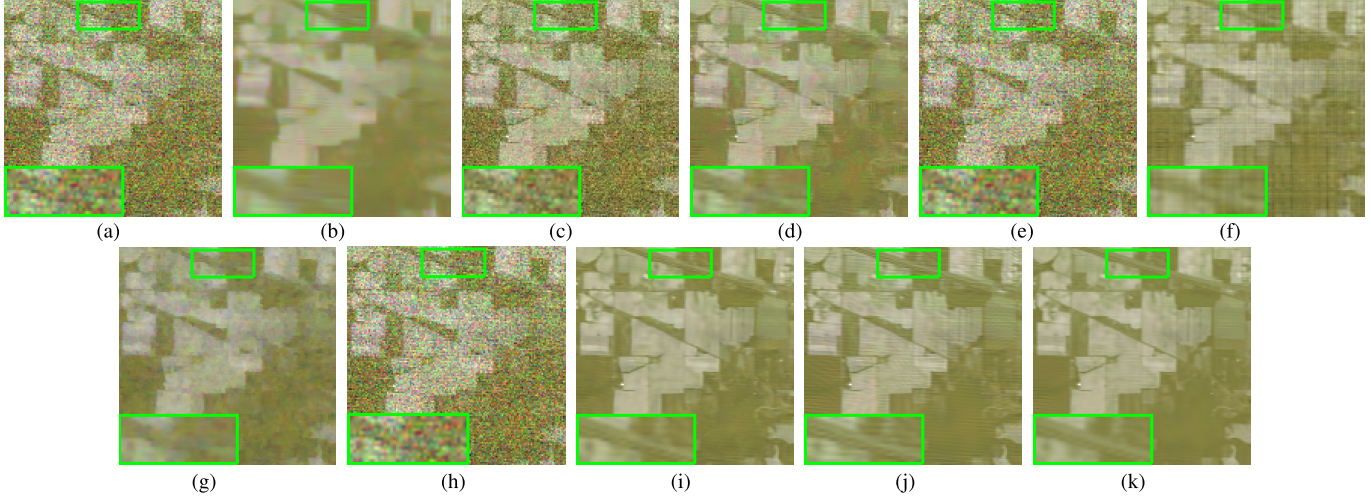

Fig. 10. Denoised results for the Indian Pines sets of real data. (a) Noisy. (b) BM3D. (c) LRMR. (d) BM4D. (e) LRTA. (f) PARAFAC. (g) LRTDTV. (h) TDL. (i) KBR. (j) LLRT. (k) NLTR.

With different noise levels, the change in the PSNR values is not very rapid except for the cases when noise variances are equal to 30 and 50 in the PaC data set, which indicates the relative robustness of the parameter P. To balance the performance and efficiency, the patch size was set to $P = 6$ in all the experiments. Moreover, the overlapped pixel was set to 1.

2) Analysis of the Number of Similar Patches K: The relationship of PSNR with similar patch numbers K is shown in Fig. 12(b). From the results, with the increase in the patch number, the PSNR values are significantly improved. When the patch number reaches 200, the PSNR results remain stable; thus, the patch number was empirically set to $K = 200$ in all the experiments.

3) Analysis of the Proximal Parameter ρ: In the NLTR solver, a proximal term was employed to guarantee algorithm stability. Fig. 12(c) shows the PSNR curves with respect to different values of ρ at four noise levels. From Fig. 12(c), it can be observed that the PSNR achieves obvious improvement with the increase in ρ. However, when ρ continued to increase, the performance worsened, and the best results were achieved with $\rho = 1$. Based on these results, the empirical setting $\rho = 1$ was used in all the experiments.

4) Analysis of TR Rank: The TR rank is an important parameter in the proposed model, and it controls the correlation of the nonlocal grouped tensor. To simplify the selection of the TR rank, it was assumed that the TR ranks of the second and third dimensions were equal, i.e., that TR rank was set as $r = [r_1, r_2, r_2]$. Fig. 13 shows the changes in the PSNR values with different TR ranks at four noise levels. From Fig. 13, it can be observed that the PSNR increases with increasing TR rank at a low noise level $\sigma = 10$. In contrast, r_2 should be smaller when the noise level is higher. The reason is that smaller TR rank can suppress noise when the noise

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 GEOSCIENCE AND REMOTE SENSING
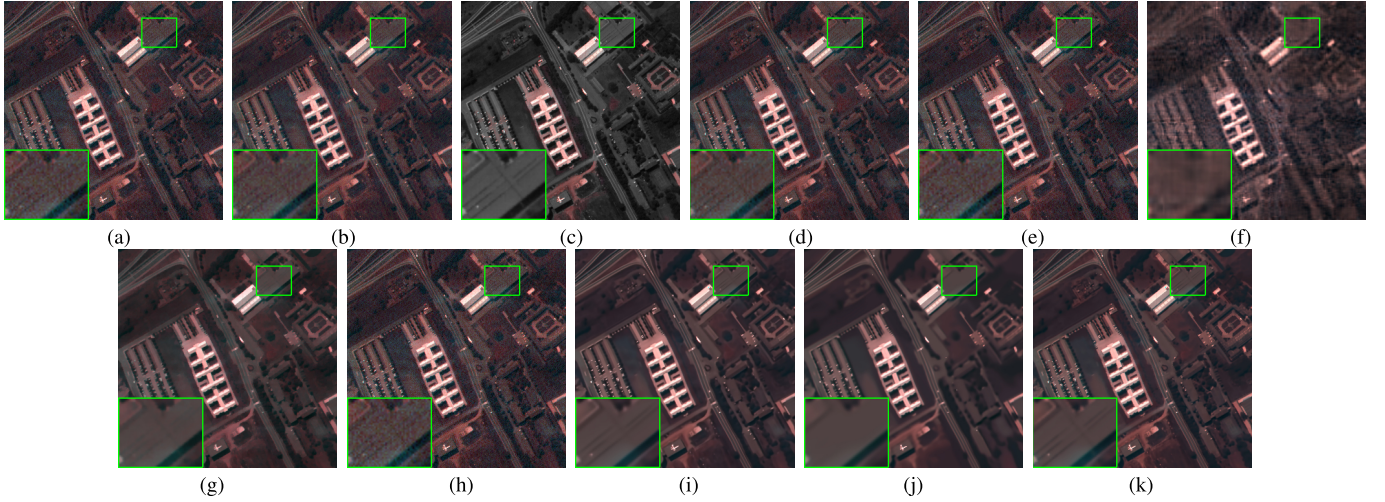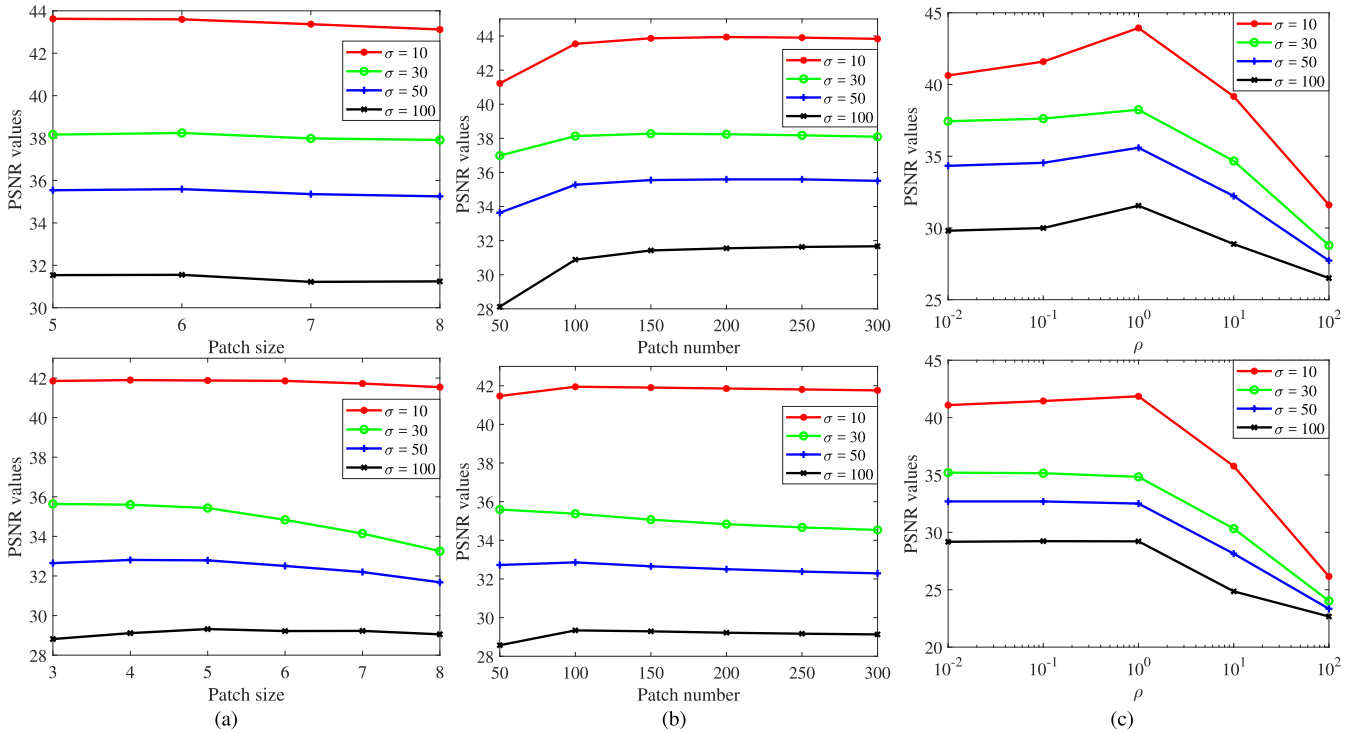

Fig. 11.   Denoised results for the Pavia University sets of real data. (a) Noisy. (b) BM3D. (c) LRMR. (d) BM4D. (e) LRTA. (f) PARAFAC. (g) LRTDTV. (h) TDL. (i) KBR. (j) LLRT. (k) NLTR.

Fig. 12.   Relationship of parameters with PSNR values at different noise levels. (a) Patch size P. (b) Number of similar patches K. (c) Proximal parameter ρ. Top: Toy data set. Bottom: PaC data set.

level is high. To balance the robustness of the TR rank and efficiency of the denoising results, the TR rank was empirically set to $r = [12, 8, 8]$ at a low noise level ($\sigma = 10$). For $\sigma = \{30, 50, 100\}$, the TR rank was set to $r = [9, 3, 3]$. In the real data, because the Indian Pines data set was seriously degraded by noise, the TR rank was set low $r = [9, 3, 3]$. For the Pavia University data set, the TR rank was set to $r = [12, 8, 8]$. In summary, the related parameters in the proposed model were relatively robust, because they could be fixed in all experiments and, thereby, obtained a better denoising result.

In our experimental results, we compared our NLTR method with many nonlocal modeling-based methods to show the superiority of TR decomposition for HSI denoising. In the next, to show the contribution of nonlocal modeling in NLTR, we compared the results of NLTR with the method that directly applies TR decomposition (without the nonlocal strategy) to the whole image. Table V lists the quantitative comparison results of NLTR with TR on the PaC data set. It can be observed that the nonlocal strategy significant improves the denoised performance, especially for high noise variance.

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

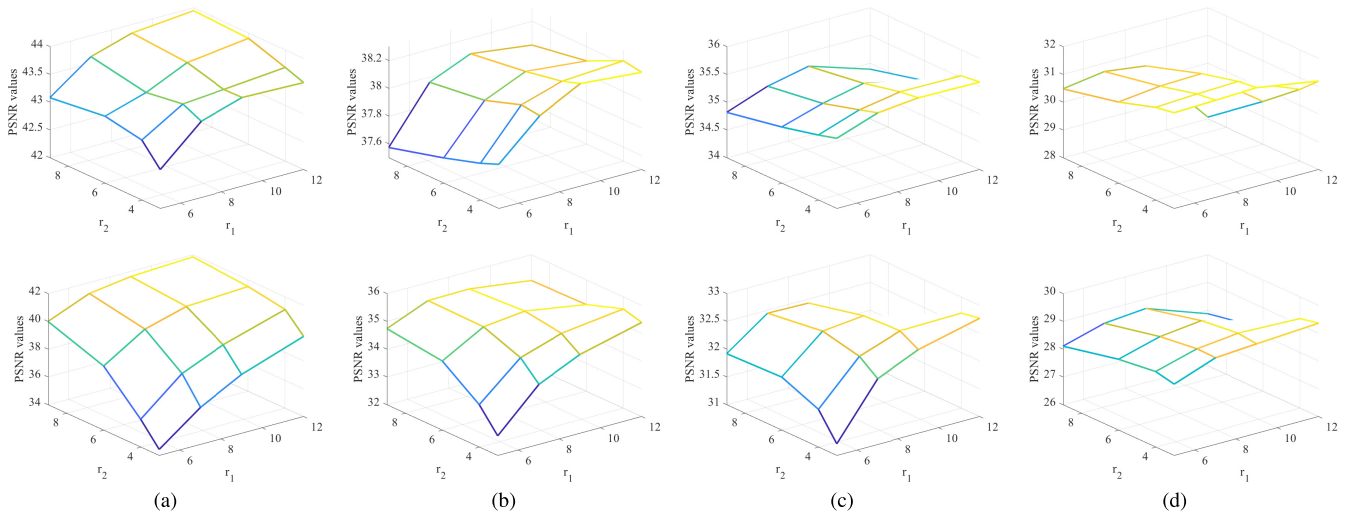CHEN *et al.*: NLTR DECOMPOSITION FOR HYPERSPECTRAL IMAGE DENOISING

13

Fig. 13. Relationship of TR rank with PSNR values at different noise levels. Top: Toy data set. Bottom: PaC data set. (a) $\sigma = 10$. (b) $\sigma = 30$. (c) $\sigma = 50$. (d) $\sigma = 100$.

TABLE V
COMPARISON OF QUANTITATIVE RESULTS OF NLTR WITH TR USED ON THE PaC DATA SET UNDER DIFFERENT NOISE VARIANCES

		PaC dataset				
σ	Index	PSNR	SSIM	FSIM	ERGAS	SAM
10	TR	39.228	0.984	0.988	37.460	0.075
	NLTR	**41.856**	**0.992**	**0.994**	**27.685**	**0.042**
30	TR	31.886	0.922	0.944	86.848	0.146
	NLTR	**34.836**	**0.963**	**0.975**	**61.934**	**0.069**
50	TR	28.794	0.857	0.901	123.655	0.177
	NLTR	**32.507**	**0.938**	**0.958**	**80.617**	**0.086**
100	TR	24.391	0.664	0.804	204.280	0.172
	NLTR	**29.218**	**0.874**	**0.917**	**117.408**	**0.113**

V. CONCLUSION

In this article, a new NLTR decomposition was proposed for HSI denoising. The TR representation was used to explore simultaneously the nonlocal self-similarity and spectral correlation in the HSI. To optimize the TR factor efficiently, the PAM algorithm was designed to optimize the model. Extensive comparative experimental testing of both simulated and real data sets, using a group of state-of-the-art HSI denoising methods, demonstrated the superiority of NLTR decomposition for HSI denoising. It is expected that this TR denoising framework will be extended in the future research for other image processing tasks.

Although the proposed model can achieve good denoising performance, there is still room for improvement. For example, automatic TR rank parameter estimation should be considered in the future, which is expected to improve the application and denoising results. In addition, the combination of a TR network with deep learning ideas [57], [58] offers a promising approach for HSI denoising.

APPENDIX

In the following, a convergence analysis of the algorithm is given briefly. First, the convergence condition of the algorithm is reviewed.

Lemma 2 ([46]): Let function $f : \mathbb{R}^n \to \mathbb{R} \cup +\infty$ be a proper lower semicontinuous function, and let $(x^j)_{j \in \mathbb{N}} \in \mathbb{R}^n$ be a sequence such that:

C1 (Sufficient decrease condition): For each $j \in \mathbb{N}$, $f(x^{j+1}) + c_1 \|x^{j+1} - x^j\|_2^2 \leq f(x^j)$ holds for a constant $c_1 \in (0, +\infty)$;

C2 (Relative error condition): For each $j \in \mathbb{N}$, $\exists\, g^{j+1} \in \partial f(x^{j+1})$ such that $\|g^{j+1}\|_2 \leq c_2 \|x^{j+1} - x^j\|_2^2$ holds for a constant $c_2 \in (0, +\infty)$;

C3 (Continuity condition): There exists a subsequence $\{x^{j_k}\}_{j \in \mathbb{N}}$ and $\bar{x} \in \mathbb{R}^n$ such that $x^{j_k} \to \bar{x}$ and $f(x^{j_k}) \to f(\bar{x})$, as $k \to +\infty$.

If the function f has the Kurdyka–Lojasiewicz (KL) property [46] at $\bar{x}$ specified in **C3**, then the sequence $(x^j)_{j \in \mathbb{N}}$ converges to $x^j \to \bar{x}$; $\bar{x}$ is a critical point of function f; and the sequence $(x^j)_{j \in \mathbb{N}}$ has a finite length, i.e., $\sum_{k=0}^{+\infty} \|x^{j+1} - x^j\| < +\infty$.

Theorem 1: Assume that the sequence $(\mathcal{G}_i)^j$ generated by the proposed algorithm is bounded, then it converges to a critical point of the objection function.

Proof: Let $f(\mathcal{G}_i) = (1/2)\|\mathcal{Y}_i - \Phi([\mathcal{G}_i])\|_F^2$; from this, it is easy to know that f is a C^1 function with Lipschitz continuous gradient. Moreover, f is a polynomial of $\mathcal{G}_i$ because of the definition of the Frobenius norm. Based on the condition that the polynomial operator is a semialgebraic function, f is semialgebraic.

According to the literature [59], a semialgebraic real-valued function f is a KL function, thereby f satisfying the KL property at $\forall\, x \in dom(f)$. In addition, the proximal operator is introduced in the solver, which is the framework of the alternating minimizing procedure [see 46, eqs. (61)–(63)]) with $B_i = \rho I$, $\rho > 0$. Thus, the iterative sequence $\mathcal{G}_i^j$ obtained by the PAM algorithm satisfies the conditions C1, C2, and C3 in Lemma 2. That is, the bounded sequences generated by the proposed algorithm converge to a critical point of f. ∎

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

14   IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING

REFERENCES

[1] 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.

[2] L. Deng, G. Vivone, W. Guo, M. D. Mura, and J. Chanussot, "A variational pansharpening approach based on reproducible kernel Hilbert space and heaviside function," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4330–4344, Sep. 2018.

[3] 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.

[4] X.-L. Zhao, F. Wang, T.-Z. Huang, M. K. Ng, and R. J. Plemmons, "Deblurring and sparse unmixing for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 4045–4058, Jul. 2013.

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

[6] 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, 2020.

[7] 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.

[8] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 2005, pp. 60–65.

[9] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1620–1630, Apr. 2013.

[10] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proc. CVPR*, Jun. 2014, pp. 2862–2869.

[11] T. Huang, W. Dong, X. Xie, G. Shi, and X. Bai, "Mixed noise removal via Laplacian scale mixture modeling and nonlocal low-rank approximation," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3171–3186, Jul. 2017.

[12] Y. Chen, W. He, N. Yokoya, and T.-Z. Huang, "Blind cloud and cloud shadow removal of multitemporal images based on total variation regularized low-rank sparsity decomposition," *ISPRS J. Photogramm. Remote Sens.*, vol. 157, pp. 93–107, Nov. 2019.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] J. Yang, Y.-Q. Zhao, J. C.-W. Chan, and S. G. Kong, "Coupled sparse denoising and unmixing with low-rank constraint for hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1818–1833, Mar. 2016.

[20] M. Ye, Y. Qian, and J. Zhou, "Multitask sparse nonnegative matrix factorization for joint spectral–spatial hyperspectral imagery denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2621–2639, May 2015.

[21] 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.

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

[23] B. Du, Z. Huang, and N. Wang, "A bandwise noise model combined with low-rank matrix factorization for hyperspectral image denoising," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1070–1081, Apr. 2018.

[24] 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.

[25] N. Renard, S. Bourennane, and J. Blanc-Talon, "Denoising and dimensionality reduction using multilinear tools for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 5, no. 2, pp. 138–142, Apr. 2008.

[26] Y. Chen, W. He, N. Yokoya, and T.-Z. Huang, "Hyperspectral image restoration using weighted group sparsity-regularized low-rank tensor decomposition," *IEEE Trans. Cybern.*, to be published. doi: 10.1109/TCYB.2019.2936042.

[27] X. Bai, F. Xu, L. Zhou, Y. Xing, L. Bai, and J. Zhou, "Nonlocal similarity based nonnegative tucker decomposition for hyperspectral image denoising," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 701–712, Mar. 2018.

[28] Y. Wang, J. Peng, Q. Zhao, D. Meng, Y. Leung, and X.-L. Zhao, "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.

[29] 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.

[30] 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.

[31] 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.

[32] Q. Xie *et al.*, "Multispectral images denoising by intrinsic tensor sparsity regularization," in *Proc. CVPR*, Jun. 2016, pp. 1692–1700.

[33] Y. Chang, L. Yan, and S. Zhong, "Hyper-Laplacian regularized unidirectional low-rank tensor recovery for multispectral image denoising," in *Proc. CVPR*, Jul. 2017, pp. 4260–4268.

[34] 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.

[35] Q. Xie, Q. Zhao, D. Meng, and Z. Xu, "Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 8, pp. 1888–1902, Aug. 2018.

[36] J. Xue, Y. Zhao, W. Liao, and S. G. Kong, "Joint spatial and spectral low-rank regularization for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 1940–1958, Apr. 2018.

[37] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "Nonlocal transform-domain filter for volumetric data denoising and reconstruction," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 119–133, Apr. 2013.

[38] Q. Zhao, G. Zhou, S. Xie, L. Zhang, and A. Cichocki, "Tensor ring decomposition," 2016, *arXiv:1606.05535*. [Online]. Available: https://arxiv.org/abs/1606.05535

[39] Q. Zhao, M. Sugiyama, L. Yuan, and A. Cichocki, "Learning efficient tensor representations with ring-structured networks," in *Proc. ICASSP*, May 2019, pp. 8608–8612.

[40] W. Wang, V. Aggarwal, and S. Aeron, "Efficient low rank tensor ring completion," in *Proc. ICCV*, Oct. 2017, pp. 5697–5705.

[41] L. Yuan, J. Cao, X. Zhao, Q. Wu, and Q. Zhao, "Higher-dimension tensor completion via low-rank tensor ring decomposition," in *Proc. APSIPA ASC*, Nov. 2018, pp. 1071–1076.

[42] L. Yuan, C. Li, D. Mandic, J. Cao, and Q. Zhao, "Tensor ring decomposition with rank minimization on latent space: An efficient approach for tensor completion," in *Proc. AAAI*, vol. 33, 2019, pp. 9151–9158.

[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] T. Uezato, M. Fauvel, and N. Dobigeon, "Hyperspectral unmixing with spectral variability using adaptive bundles and double sparsity," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3980–3992, Jun. 2019.

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.*: NLTR DECOMPOSITION FOR HYPERSPECTRAL IMAGE DENOISING

15

[46] H. Attouch, J. Bolte, and B. F. Svaiter, "Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods," *Math. Program.*, vol. 137, nos. 1–2, pp. 91–129, 2013.

[47] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126–1153, 2013.

[48] T.-X. Jiang, T.-Z. Huang, X.-L. Zhao, T.-Y. Ji, and L.-J. Deng, "Matrix factorization for low-rank tensor completion using framelet prior," *Inf. Sci.*, vol. 436, pp. 403–417, Apr. 2018.

[49] Y.-B. Zheng, T.-Z. Huang, T.-Y. Ji, X.-L. Zhao, T.-X. Jiang, and T.-H. Ma, "Low-rank tensor completion via smooth matrix factorization," *Appl. Math. Model.*, vol. 70, pp. 677–695, Jun. 2019.

[50] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 127–239, Jan. 2014.

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

[52] 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.

[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] L. Wald, *Data Fusion: Definitions and Architectures: Fusion of Images of Different Spatial Resolutions*. Paris, France: Presses des MINES, 2002.

[56] 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.

[57] 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.

[58] 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.

[59] J. Bolte, A. Daniilidis, A. Lewis, and M. Shiota, "Clarke subgradients of stratifiable functions," *SIAM J. Optim.*, vol. 18, no. 2, pp. 556–572, 2007.
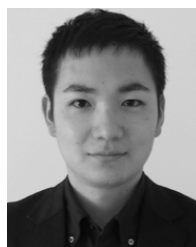
Naoto Yokoya (S'10–M'13) received the M.Eng. and Ph.D. degrees in aerospace engineering from the University of Tokyo, Tokyo, Japan, in 2010 and 2013, respectively.

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

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

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, in 1986, 1992, and 2001, respectively.

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

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

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 (UESTC), Chengdu, China.

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

Wei He (S'14–M'17) received the B.S. degree from the School of Mathematics and Statistics and the Ph.D. degree in surveying, mapping and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan, China, in 2012 and 2017, respectively.

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

Xi-Le Zhao received the M.S. and Ph.D. degrees from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2009 and 2012, respectively.

He is currently a Professor with the School of Mathematical Sciences, UESTC. His research interests include image processing, computer vision, and machine learning.