Multispectral image denoising using sparse and graph Laplacian Tucker decomposition

Xiaoce Wu¹, Bingyin Zhou¹ (✉), Qingyun Ren¹, and Wei Guo¹

© The Author(s) 2020.

Abstract Multispectral image denoising is a basic problem whose results affect subsequent processes such as target detection and classification. Numerous approaches have been proposed, but there are still many challenges, particularly in using prior knowledge of multispectral images, which is crucial for solving the ill-posed problem of noise removal. This paper considers both non-local self-similarity in space and global correlation in spectrum. We propose a novel low-rank Tucker decomposition model for removing the noise, in which sparse and graph Laplacian regularization terms are employed to encode this prior knowledge. It can jointly learn a sparse and low-rank representation while preserving the local geometrical structure between spectral bands, so as to better capture simultaneously the correlation in spatial and spectral directions. We adopt the alternating direction method of multipliers to solve the resulting problem. Experiments demonstrate that the proposed method outperforms the state-of-the-art, such as cube-based and tensor-based methods, both quantitatively and qualitatively.

Keywords denoising; tensor decomposition; sparsity; graph Laplacian; multispectral image

1 Introduction

A multispectral image (MSI) contains dozens of bands, each of which is captured over a specific wavelength range of the electromagnetic spectrum. It provides rich spectral-spatial information for improved performance of applications in remote sensing, such as target detection [1], classification [2], and tracking [3]. However, inevitably there are

various kinds of noises which degrade visual quality, and affect the results obtained in those applications. MSI denoising is thus useful as a basic preprocessing step which can improve performance of subsequent processes. In recent years, it has attracted more and more attention in the fields of computer vision and remote sensing.

A great many MSI denoising approaches have been proposed to date [4]. The spatial-spectral structure of an MSI corresponds to a 1D signal at each spatial point, and a grayscale image in each spectral band. Hence, straightforward denoising employs conventional methods for 1D signals and grayscale images, such as wavelet domain soft thresholding [5], K -SVD [6], and block-matching and 3D filtering (BM3D) [7]. However, those methods only take into account spectral correlation or spatial correlation, and thus satisfactory results cannot be generally achieved. To overcome this issue, various methods aimed at making full use of spatial-spectral correlation have been proposed. Some examples include the use of principal component analysis [8], multidimensional Wiener filtering [9], and tensor decompositions [10–12].

Because MSI denoising is inherently ill-posed, formulating and exploiting prior knowledge plays a central role in addressing the problem. Non-local spatial self-similarity and global correlation in spectrum, which have been shown to be two very reasonable priors [13–15], are commonly used. Non-local self-similarity refers to a phenomenon when observing an MSI from a perspective of small full-band patches: each full-band patch is similar to many other, non-local, ones. Global correlation in spectrum means that there is a large amount of spectral redundancy in an MSI: different bands' images are generally highly correlated.

¹ Key Laboratory of Augmented Reality, School of Mathematical Science, Hebei Normal University, Shijiazhuang 050024, China. E-mail: zhoubingyin@163.com (✉).

Manuscript received: 2020-03-29; accepted: 2020-04-26

In the state-of-the-art approach proposed in Ref. [16], to use these two kinds of priors, a Kronecker basis representation (KBR) based tensor sparsity measure is introduced and applied to tensors formed from non-local similar full-band patches. However, it fails to consider the geometrical structure linking spectral bands: the set of clean band images lies on a low-dimensional manifold embedded in image space. Thus, as the results of band selection in Ref. [17] show, there are possibilities to further investigate spectral correlations using this geometrical structure, to obtain better denoising results.

This paper aims to take a further step in effectively utilizing the priors of non-local self-similarity and spectral correlation. As in Ref. [16], we build tensors from non-local similar full-band patches to naturally form representations based on those two priors. The low-rank Tucker decomposition is adopted to capture correlation simultaneously in all dimensions of the patch tensor. Furthermore, to fully identify the two kinds of priors, the band-wise local geometrical structure is formulated and incorporated as well as sparse regularization. Our main contributions are:

- A sparse and graph Laplacian regularized low-rank Tucker decomposition model for removing noise. We use weighted graphs to characterize the intrinsic local geometrical structure of spectral bands. As similar bands should have similar low-dimensional representations, graph embedding is used to jointly learn a sparse low-rank representation while preserving the graph structure.
- An algorithm based on this model and the altering direction method of multipliers (ADMM) for removing noise.
- An evaluation of this approach based on a benchmark MSI dataset with various noise levels. Detailed analyses such as the balance of noise removal in different spectral bands are carried out. The results show that our method outperforms state-of-the-art methods, both quantitatively and qualitatively.

The rest of the paper is organized as follows. In Sections 2 and 3, we respectively introduce useful tensor concepts and give a brief review of related work. Section 4 investigates how to encode the priors, and proposes a sparse, graph Laplacian regularized, low-rank Tucker decomposition model for

MSI denoising. A denoising algorithm based on the alternating direction method of multipliers employing this model is also given. In Section 5, experiments show quantitative and qualitative evaluation of the proposed method. Finally, we conclude the paper in Section 6.

2 Tensor background

In this section, we briefly describe some useful tensor concepts. Readers are referred to Refs. [18, 19] for details and applications.

First, we explain the notation used in this paper. We denote scalars by non-bold letters, vectors by bold lowercase letters, matrices by bold uppercase letters, and higher-order tensors by bold calligraphy letters. The i th column of a matrix $\mathbf{A}$ is denoted by $\mathbf{a}_i$ or $\mathbf{A}(:, i)$, and the i th row by $\mathbf{A}(i, :)$. We use $[N]$ to denote the set $\{1, \dots, N\}$ for $N \in \mathbb{N}_+$.

Tensors are elements of a tensor product of vector spaces, and often assimilated to corresponding array representations while fixing bases on the vector spaces. Specifically, a real N th-order $(I_1, \dots, I_N)$ -dimensional tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times \dots \times I_N}$ is an $I_1 \times \dots \times I_N$ array:

$$\mathcal{A} = (a_{i_1 \dots i_N}), \quad a_{i_1 \dots i_N} \in \mathbb{R}$$

where $i_n \in [I_n]$ for $n \in [N]$. Scalars, vectors, and matrices are 0th-order, 1st-order, and 2nd-order tensors respectively. Tensors of order ≥ 3 are called higher-order tensors, and are natural representations of multidimensional data preserving their native structure. For example, an MSI can be naturally treated as a 3rd-order tensor, in which the spatial-spectral structure is preserved.

Let $\mathcal{A} \in \mathbb{R}^{I_1 \times \dots \times I_N}$. A mode- n vector of tensor $\mathcal{A}$ is an I_n -dimensional vector defined by fixing all indices but i_n ; the mode- n matricization, denoted $\mathbf{A}_{(n)}$, is an $I_n \times (I_1 \dots I_{n-1} I_{n+1} \dots I_N)$ matrix formed by arranging all mode- n vectors as its columns. For instance, the mode-2 matricization of a matrix is its transpose.

The scalar product of tensors $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{I_1 \times \dots \times I_N}$ is defined as

$$\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i_1=1}^{I_1} \dots \sum_{i_N=1}^{I_N} a_{i_1 \dots i_N} b_{i_1 \dots i_N} \quad (1)$$

It is clear that the scalar product of tensors can be expressed in matrix form as

$$\langle \mathcal{A}, \mathcal{B} \rangle = \text{tr} \left(\mathbf{A}_{(n)}^T \mathbf{B}_{(n)} \right), \quad n \in [N]$$

in which $\text{tr}(\cdot)$ denotes the trace operator of matrix.

The Frobenius norm of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ is given by

$$\|\mathcal{A}\|_F = \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle} \quad (2)$$

The n -mode product of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ with a matrix $\mathbf{U} \in \mathbb{R}^{J_n \times I_n}$, denoted $\mathcal{A} \times_n \mathbf{U}$, is a tensor $\mathcal{B} \in \mathbb{R}^{I_1 \times \cdots \times I_{n-1} \times J_n \times I_{n+1} \times \cdots \times I_N}$ defined entry-wise by

$$b_{i_1 \cdots i_{n-1} j_n i_{n+1} \cdots i_N} = \sum_{i_n=1}^{I_n} a_{i_1 \cdots i_N} u_{j_n i_n} \quad (3)$$

It can be matricized as

$$\mathbf{B}_{(n)} = \mathbf{U} \mathbf{A}_{(n)}, \quad n \in [N] \quad (4)$$

Each mode- n vector of $\mathcal{A}$ is multiplied by $\mathbf{U}$.

A tensor $\mathcal{S} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ is said to be rank-1 if it can be expressed as the outer product of N vectors:

$$\mathcal{S} = \mathbf{v}^{(1)} \circ \cdots \circ \mathbf{v}^{(N)} \quad (5)$$

where $\mathbf{v}^{(n)} \in \mathbb{R}^{I_n}$ for $n \in [N]$.

Let $\mathcal{A}$ be an arbitrary $I_1 \times \cdots \times I_N$ tensor. The rank of $\mathcal{A}$, denoted $\text{rank}(\mathcal{A})$, is defined as the minimal number of rank-1 tensors that yield $\mathcal{A}$ as their sum; the n -rank $\text{rank}_n(\mathcal{A})$ is defined as the column rank of $\mathbf{A}_{(n)}$; the multilinear rank is consequently defined as the N -tuple $(R_1, \dots, R_N)$, in which $R_n = \text{rank}_n(\mathcal{A})$. Note that the values of $R_1, \dots, R_N$ and $R = \text{rank}(\mathcal{A})$ can all be different for higher order tensors.

Using the concepts of tensors, the singular value decomposition (SVD) of a matrix $\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T$ can be written as $\mathbf{A} = \mathbf{\Sigma} \times_1 \mathbf{U} \times_2 \mathbf{V}$. This means that a 2nd-order tensor $\mathbf{A}$ can be decomposed into a diagonal core tensor $\mathbf{\Sigma}$ multiplied by orthonormal matrices $\mathbf{U}$ and $\mathbf{V}$ along its 1-mode and 2-mode respectively. Tensor decompositions are useful in multilinear algebra. They are higher-order generalizations of the matrix SVD. However, a decomposition that satisfies both properties, namely orthogonality of the factor matrices and diagonality of the core tensor, does not exist in general. There are two basic models that treat a tensor as a global quantity.

One basic model is the CP decomposition proposed by Hitchcock [20]. It factorizes a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ into a linear combination of R rank-1 tensors:

$$\mathcal{A} = \sum_{r=1}^R \lambda_r \left(\mathbf{v}_r^{(1)} \circ \cdots \circ \mathbf{v}_r^{(N)} \right) \quad (6)$$

where $\lambda_r \in \mathbb{R}$ and $\mathbf{v}_r^{(n)} \in \mathbb{R}^{I_n}$ are normalized vectors for $r \in [R]$ and $n \in [N]$.

The other basic model is Tucker decomposition [21]. It factorizes a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ into a core tensor $\mathcal{G} \in \mathbb{R}^{R_1 \times \cdots \times R_N}$ multiplied by a matrix $\mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times R_n}$ along each n -mode:

$$\begin{aligned} \mathcal{A} &= \mathcal{G} \times_1 \mathbf{U}^{(1)} \times_2 \cdots \times_N \mathbf{U}^{(N)} \\ &= \sum_{r_1=1}^{R_1} \cdots \sum_{r_N=1}^{R_N} y_{r_1 \cdots r_N} \mathbf{u}_{r_1}^{(1)} \circ \cdots \circ \mathbf{u}_{r_N}^{(N)} \end{aligned} \quad (7)$$

Here, the factor matrices $\mathbf{U}^{(1)}, \dots, \mathbf{U}^{(N)}$ are usually column orthonormal. This decomposition can be expressed in matrix form as

$$\begin{aligned} \mathbf{A}_{(n)} &= \mathbf{U}^{(n)} \mathbf{G}_{(n)} \left(\mathbf{U}^{(N)} \otimes \cdots \otimes \mathbf{U}^{(n+1)} \right. \\ &\quad \left. \otimes \mathbf{U}^{(n-1)} \otimes \cdots \otimes \mathbf{U}^{(1)} \right)^T, \quad n \in [N] \end{aligned}$$

The symbol $\otimes$ represents the Kronecker product of matrices.

3 Related work

Many methods have been proposed for MSI denoising. This section briefly reviews the relevant literature.

3.1 2D and extended 2D approaches

As mentioned earlier, grayscale image denoising approaches, such as non-local means (NLM) [22], K -SVD [6], and BM3D [7], can be directly used to reduce noise in each spectral band separately. However, an MSI is not just a collection of grayscale images: there is significant correlation between the spectral bands. Hence, these methods generally cannot achieve satisfactory results. In Ref. [23], a spectral-spatial kernel method is proposed. It can maintain spectral correlation and simultaneously match the original structure between two spatial dimensions.

Another reasonable idea is to extend patch-based 2D denoising methods to 3D by building small cubes from an MSI. An adaptive non-local means (ANLM3D) method is proposed in Ref. [24]; the state-of-the-art method in this line is BM4D proposed in Ref. [13]. It utilizes 3D non-local self-similarity to collaboratively remove noise in similar cubes. Such methods still have not reached the full potential for noise removal, since the global spectral correlation prior is neglected.

3.2 Tensor-based approaches

Tensor decompositions are promising tools to process multidimensional data, preserving the native form of the data. Recently, the mathematical foundation of

tensor decompositions has rapidly developed as well as tensor-based data processing techniques [18, 19]. An MSI can be naturally represented by a 3rd-order tensor, and many denoising methods applying various tensor decompositions have been proposed to date.

In Ref. [12], the CP decomposition is used. Based on the Tucker decomposition, a multidimensional Wiener filtering method is presented in Ref. [9]. A multilinear low-rank approximation method is proposed in Ref. [10]. To avoid explicitly estimating the rank parameters, a sparse regularization on the core tensor is induced in Ref. [25]. The recently proposed tensor singular value decomposition (t-SVD) [26] is adopted in Ref. [27] to achieve an effective representation for the MSI, and thus effective recovery. The advantage of these methods is that they jointly capture the correlation in spatial and spectral directions, but the non-local self-similarity prior is not well utilized.

The tensor dictionary learning (TDL) [14] method takes the two kinds of priors mentioned earlier into account, but it just coarsely encodes the non-local self-similarity prior using a relatively small number of full-band clusters. The state-of-the-art method in this line was proposed in Ref. [16], in which a KBR based sparsity measure is introduced and applied to tensors formed by non-local similar full-band patches. However, it does not consider the geometrical structure linking spectral bands, i.e., that the set of clean band images resides on a low-dimensional manifold. There are possibilities to further investigate spectral correlation by using the geometrical structure, thereby improving the denoising performance, as the results of band selection show in Ref. [17].

4 Sparse and graph regularized low-rank Tucker decomposition

We treat an MSI as an $H \times W \times S$ tensor, where H , W , and S are the height, width, and number of spectral bands respectively. Driven by the idea of using prior knowledge to improve the results of denoising, this paper investigates the priors of non-local self-similarity and spectral correlation based on a patch framework. We propose a low-rank Tucker decomposition model with sparse and graph Laplacian regularization, in which a sparse low-rank representation incorporating the local geometrical

information linking spectral bands can be jointly learned. Its suitability for combining MSI correlation simultaneously along spatial and spectral directions produces better results. Figure 1 illustrates the idea of our method.

4.1 Problem formulation

Typically, an MSI is degraded by additive Gaussian noise. Hence, we consider the noise degradation model formulated as

$$\mathbf{Y} = \mathbf{X} + \mathbf{\varepsilon} \quad (8)$$

where $\mathbf{Y}$ is the noisy MSI, $\mathbf{X}$ is the clean MSI, and $\mathbf{\varepsilon}$ is the Gaussian noise. The problem of MSI denoising, that is recovering $\mathbf{X}$ from $\mathbf{Y}$, is inherently ill-posed, and prior knowledge is needed to address the problem.

This paper employs both non-local self-similarity and spectral correlation priors at the same time, as do other state-of-the-art methods [13, 14, 16]. For each $h \times w \times S$ full-band patch $\mathcal{P}^{(i,1)}$ in $\mathbf{Y}$, a group of $M - 1$ non-local patches similar to $\mathcal{P}^{(i,1)}$, denoted $\mathcal{P}^{(i,2)}, \dots, \mathcal{P}^{(i,M)}$, is extracted from $\mathbf{Y}$ by block matching [7]. Then, an $S \times wh \times M$ patch tensor $\mathbf{Y}^{(i)}$ is built up by stacking each mode-3 matricization of those M patches:

$$\mathbf{Y}^{(i)}(:, :, m) = \mathbf{P}_{(3)}^{(i,m)}, \quad m \in [M]$$

The recovery of the underlying clean patch tensor $\mathbf{X}^{(i)}$ for $\mathbf{Y}^{(i)}$ is formulated as an optimization problem:

$$\min_{\mathbf{X}^{(i)}} R(\mathbf{X}^{(i)}) + \frac{\lambda}{2} \left\| \mathbf{Y}^{(i)} - \mathbf{X}^{(i)} \right\|_{\text{F}}^2 \quad (9)$$

in which $R(\mathbf{X}^{(i)})$ is the regularization term identifying prior knowledge about clean patch tensors, and $\lambda > 0$ is a constant controlling the tradeoff between the two terms. Finally, the overall MSI can be estimated by aggregating all restored patches.

For simplicity and without confusion, we hereafter omit the superscripts for patch tensors. The Tucker decomposition is used to aggregate the correlation simultaneously in all different dimensions:

$$\mathbf{X} = \mathbf{G} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W} \quad (10)$$

in which $\mathbf{G} \in \mathbb{R}^{R_1 \times R_2 \times R_3}$ is a core tensor, and $\mathbf{U} \in \mathbb{R}^{S \times R_1}$, $\mathbf{V} \in \mathbb{R}^{wh \times R_2}$, and $\mathbf{W} \in \mathbb{R}^{M \times R_3}$ are column orthonormal factor matrices.

The Tucker decomposition is not unique [18]. This opens the door to placing constraints on the core tensor to improve the uniqueness, thereby improving the performance of noise removal. Next, a formula for the regularization term in model (9) will be investigated specifically.

Fig. 1 Concept of our method. Superscripts i in full-band patches, and (i) in patch tensors are omitted for simplicity. We investigate the priors of non-local self-similarity and spectral correlation based on a patch framework. Non-local similar full-band patches $\mathcal{P}^{(2)}, \dots, \mathcal{P}^{(M)}$ are extracted for a selected full-band patch $\mathcal{P}^{(1)}$, and a noisy patch tensor $\mathcal{Y}$ delivering the two priors is built by matricizing and stacking those patches. We use Tucker decomposition to capture the correlation simultaneously in all different directions. In this representation, the patch tensor $\mathcal{X}$ of a clean MSI has the properties of multilinear low-rank and a sparse core tensor. Furthermore, the local geometrical structure linking spectral bands is also considered. We use a weighted graph, reconstructed using the noisy patch tensor $\mathcal{Y}$ or spectral priors, to characterize the intrinsic local structure. Graph embedding is utilized to jointly learn a sparse low-rank representation while preserving the graph structure.

4.2 Sparse and low-rank regularization

By construction, intuitively, the patch tensors should be low-rank in each mode, so

$$R_1 < S, \quad R_2 < wh, \quad \text{and} \quad R_3 < M$$

Hence, a direct way is just to estimate the rank parameters and then make a multilinear low-rank approximation. But it is difficult to obtain accurate rank estimates. In Ref. [25], the authors induced a sparse regularization on the core tensor, such as $\|\mathcal{G}\|_1$, instead of explicitly estimating those rank parameters.

Another natural way is to introduce a low-rank part into the regularization term in the model in Eq. (9). However, there are many ways to formulate the low rank of tensors. The recently defined tensor trace norm [28]:

$$\|\mathcal{A}\|_* = \sum_{n=1}^N \omega_n \|\mathcal{A}_{(n)}\|_*, \quad \text{with } \omega_n \geq 0, \sum_{n=1}^N \omega_n = 1$$

is widely used in denoising methods with different settings of ω_n . For example, equal weights are used

in Refs. [29, 30]: $\omega_n = \frac{1}{N}, n \in [N]$. In Ref. [31], the product of trace norm of each mode- n matricization is adopted, which is equivalent to setting weights adaptively, since

$$\prod_{n=1}^N \|\mathcal{A}_{(n)}\|_* = \sum_{n=1}^N \omega_n \|\mathcal{A}_{(n)}\|_*,$$

$$\text{with } \omega_n = \frac{1}{N} \prod_{i=1, i \neq n}^N \|\mathcal{A}_{(i)}\|_*$$

It is clear that the mode- n subspace with lower rank has a larger weight, and vice versa. Furthermore, based on the rank-product framework, the method proposed in Ref. [15] also considers fine-grained sparsity inside the core tensor, allowing it to well describe the detailed information in the data.

We follow Ref. [15] in this work, and formulate the sparse and low-rank regularization term in Eq. (9) as

$$R_1(\mathcal{X}) = \omega \|\mathcal{G}\|_0 + (1 - \omega) \prod_{k=1}^3 \text{rank}_k(\mathcal{X}) \quad (11)$$

where $\omega \in (0, 1)$ is a constant. As mentioned

in Refs. [15, 16], the regularization constrains the volume of the nonzero top-left block in the core tensor, as well as the number of Kronecker bases used for the Tucker representation. By construction of the patch tensors, it is easy to see that the singular values in the non-local self-similarity model generally decay much faster than those in the spectral and spatial models: $R_3 \ll R_1, R_2$. Thus, this regularization has a larger penalty for R_3 than for R_1 and R_2 , which is in accord with our prior knowledge underlying patch tensors.

4.3 Graph Laplacian regularization

As demonstrated in Refs. [15, 16], the sparse and low-rank regularization term Eq. (11) can efficiently encode prior knowledge concerning both non-local spatial self-similarity and global correlation in spectrum. However, one drawback is that it does not consider the geometrical structure linking spectral bands, namely that the set of clean band images lie on a low-dimensional manifold embedded in image space. Further improvement is possible by using the geometrical structure, as the results of band selection in Ref. [17] show.

Motivated by recently developed graph embeddings for manifold learning and regularization with tensor decompositions [32–34], this paper uses weighted graphs to characterize the intrinsic local geometrical structure linking spectral bands, and induces the graph Laplacian regularization aiming to jointly learn a sparse and low-rank representation while preserving the structure of the graph. Specifically, we rewrite Eq. (10) in matrix form as

$$\mathbf{X}_{(1)} = \mathbf{U}\mathbf{G}_{(1)}(\mathbf{W} \otimes \mathbf{V})^T$$

This implies that the i th band $\mathbf{X}_{(1)}(i, :) \in \mathbb{R}^{whM}$ has a low-dimensional representation $\mathbf{U}(i, :) \in \mathbb{R}^{R_1}$ in the basis consisting of all rows of $\mathbf{G}_{(1)}(\mathbf{W} \otimes \mathbf{V})^T$. The idea is that similar bands should have similar representations.

In the framework of graph embedding, an edge-weighted graph G with all bands as its vertices is built to model the local geometrical structure of the spectral bands. Let $\mathbf{C} \in \mathbb{R}^{S \times S}$ be the weight matrix of graph G . Entry c_{ij} measures the similarity between the i th and j th bands. The goal is to learn an optimal low-dimensional representation $\mathbf{U}$ for the vertices of graph G , which optimally characterizes their similarities. This can be achieved by introducing

a novel regularization term:

$$\begin{aligned} R_2(\mathbf{X}) &= \eta \sum_{i=1}^S \sum_{j=1}^S \|\mathbf{U}(i, :) - \mathbf{U}(j, :)\|^2 c_{ij} \\ &= \eta \operatorname{tr}(\mathbf{U}^T \mathbf{L} \mathbf{U}) \end{aligned} \quad (12)$$

where $\eta > 0$ is a constant, and $\mathbf{L} \in \mathbb{R}^{S \times S}$ is the graph Laplacian defined by

$$\mathbf{L} = \mathbf{D} - \mathbf{C}$$

with $\mathbf{D} = \operatorname{diag}(d_1, \dots, d_S)$ and $d_i = \sum_{j=1}^S c_{ij}$.

4.4 Proposed model

By adding the regularization terms in Eqs. (11) and (12) to the model in Eq. (9), our final model for MSI denoising is given by

$$\begin{aligned} \min_{\mathbf{G}, \mathbf{U}, \mathbf{V}, \mathbf{W}} \quad & \|\mathbf{G}\|_0 + \alpha \prod_{n=1}^3 \operatorname{rank}_n(\mathbf{X}) \\ & + \beta \operatorname{tr}(\mathbf{U}^T \mathbf{L} \mathbf{U}) + \frac{\gamma}{2} \|\mathbf{Y} - \mathbf{X}\|_F^2, \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{G} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W}, \\ & \mathbf{U}^T \mathbf{U} = \mathbf{I}, \quad \mathbf{V}^T \mathbf{V} = \mathbf{I}, \quad \mathbf{W}^T \mathbf{W} = \mathbf{I} \end{aligned} \quad (13)$$

where $\alpha = (1 - \omega)/\omega$, $\beta = \eta/\omega$, and $\gamma = \lambda/\omega$.

It is obvious that Eq. (13) is a non-convex optimization problem, since the l_0 and rank terms are all non-convex. As in Ref. [16], we use a log-sum form to approximate the l_0 norm of tensors and the rank of matrices. Let:

$$\begin{aligned} L_1(\mathbf{A}) &= \sum_{i_1, \dots, i_N} \frac{\log(|a_{i_1 \dots i_N}| + \varepsilon) - \log(\varepsilon)}{-\log(\varepsilon)} \\ L_*(\mathbf{A}) &= \sum_i \frac{\log(\sigma_i(\mathbf{A}) + \varepsilon) - \log(\varepsilon)}{-\log(\varepsilon)} \end{aligned}$$

in which ε is a small positive constant, and $\sigma_i(\mathbf{A})$ is the i th singular value of matrix $\mathbf{A}$. This leads to the following optimization problem:

$$\begin{aligned} \min_{\mathbf{G}, \mathbf{U}, \mathbf{V}, \mathbf{W}} \quad & L_1(\mathbf{G}) + \alpha \prod_{n=1}^3 L_*(\mathbf{X}_{(n)}) \\ & + \beta \operatorname{tr}(\mathbf{U}^T \mathbf{L} \mathbf{U}) + \frac{\gamma}{2} \|\mathbf{Y} - \mathbf{X}\|_F^2, \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{G} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W}, \\ & \mathbf{U}^T \mathbf{U} = \mathbf{I}, \quad \mathbf{V}^T \mathbf{V} = \mathbf{I}, \quad \mathbf{W}^T \mathbf{W} = \mathbf{I} \end{aligned} \quad (14)$$

4.5 Model optimization

We adopt the ADMM method [35] to solve the problem in Eq. (14), since it is so hard to estimate the multiple unknowns directly. By introducing

three auxiliary tensors $\mathcal{Z}^{(1)}, \mathcal{Z}^{(2)}, \mathcal{Z}^{(3)} \in \mathbb{R}^{S \times wh \times M}$, Eq. (14) can be equivalently reformulated as

$$\begin{aligned} \min_{\mathcal{G}, \mathbf{U}, \mathbf{V}, \mathbf{W}} \quad & L_1(\mathcal{G}) + \alpha \prod_{n=1}^3 L_*(\mathcal{Z}_{(n)}^{(n)}) + \beta \operatorname{tr}(\mathbf{U}^T \mathbf{L} \mathbf{U}) \\ & + \frac{\gamma}{2} \|\mathcal{Y} - \mathcal{G} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W}\|_F^2, \\ \text{s.t.} \quad & \mathbf{U}^T \mathbf{U} = \mathbf{I}, \quad \mathbf{V}^T \mathbf{V} = \mathbf{I}, \quad \mathbf{W}^T \mathbf{W} = \mathbf{I}, \\ & \mathcal{G} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W} - \mathcal{Z}^{(n)} = 0, \quad n \in [3] \end{aligned}$$

Then, the augmented Lagrangian function can be defined as

$$\begin{aligned} L(\mathcal{Z}^{(1)}, \mathcal{Z}^{(2)}, \mathcal{Z}^{(3)}, \mathcal{G}, \mathbf{U}, \mathbf{V}, \mathbf{W}) = & L_1(\mathcal{G}) \\ & + \alpha \prod_{n=1}^3 L_*(\mathcal{Z}_{(n)}^{(n)}) + \beta \operatorname{tr}(\mathbf{U}^T \mathbf{L} \mathbf{U}) \\ & + \frac{\gamma}{2} \|\mathcal{Y} - \mathcal{G} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W}\|_F^2 \\ & + \sum_{n=1}^3 \left\langle \mathcal{G} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W} - \mathcal{Z}^{(n)}, \mathcal{T}^{(n)} \right\rangle \\ & + \sum_{n=1}^3 \frac{\mu}{2} \left\| \mathcal{G} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W} - \mathcal{Z}^{(n)} \right\|_F^2 \end{aligned}$$

where $\mu > 0$ is a constant, $\mathcal{T}^{(1)}, \mathcal{T}^{(2)},$ and $\mathcal{T}^{(3)}$ are Lagrange multipliers, and $\mathbf{U}, \mathbf{V},$ and $\mathbf{W}$ are column orthonormal matrices. In the ADMM framework, the problem can now be solved by solving the following sub-problems. Before starting, we first define:

$$\mathcal{P} = \frac{1}{\gamma + 3\mu} (\gamma \mathcal{Y} + \sum_{i=1}^3 (\mu \mathcal{Z}^{(i)} - \mathcal{T}^{(i)}))$$

(1) Updating $\mathbf{U}, \mathbf{V},$ and $\mathbf{W}$. There is a closed-form solution to the problem of optimizing only $\mathbf{V}$ or $\mathbf{W}$ while keeping the other variables fixed. For example, $\mathbf{V}$ can be updated by solving

$$\min_{\mathbf{V}^T \mathbf{V} = \mathbf{I}} \|\mathcal{G} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W} - \mathcal{P}\|_F^2 \quad (15)$$

Since $\mathbf{U}, \mathbf{V},$ and $\mathbf{W}$ are column orthonormal, the problem can be rewritten as

$$\max_{\mathbf{V}^T \mathbf{V} = \mathbf{I}} \operatorname{tr}(\mathbf{B}^T \mathbf{V}) \quad (16)$$

where $\mathbf{B} = \mathbf{P}_{(2)}(\mathbf{W} \otimes \mathbf{U}) \mathbf{G}_{(2)}^T$. Thus, we can obtain a closed-form solution by using von Neumann's trace inequality [36].

With other variables fixed, $\mathbf{U}$ can be updated by solving

$$\begin{aligned} \min_{\mathbf{U}^T \mathbf{U} = \mathbf{I}} \quad & \|\mathcal{G} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W} - \mathcal{P}\|_F^2 \\ & + \beta \operatorname{tr}(\mathbf{U}^T \mathbf{L} \mathbf{U}) \end{aligned} \quad (17)$$

This is equivalent to

$$\min_{\mathbf{U}^T \mathbf{U} = \mathbf{I}} \frac{1}{2} \operatorname{tr}(\mathbf{U}^T \mathbf{A} \mathbf{U}) + \operatorname{tr}(\mathbf{B}^T \mathbf{U}) \quad (18)$$

where $\mathbf{A} = \beta \mathbf{L}$ and $\mathbf{B} = -\mathbf{P}_{(1)}(\mathbf{W} \otimes \mathbf{V}) \mathbf{G}_{(1)}^T$. When $\mathbf{B} \neq \mathbf{0}$, this problem is difficult to solve [37]. Here we use the conjugate gradient algorithm on a Stiefel manifold [38] to solve it. Note that this is the most time consuming step in the algorithm.

(2) Updating $\mathcal{G}$. With other variables fixed, we update $\mathcal{G}$ by solving

$$\min_{\mathcal{G}} \quad \rho L_1(\mathcal{G}) + \frac{1}{2} \|\mathcal{G} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W} - \mathcal{P}\|_F^2 \quad (19)$$

where $\rho = 1/(\gamma + 3\mu)$. Since $\mathbf{U}, \mathbf{V},$ and $\mathbf{W}$ are column orthonormal, the problem can be rewritten as

$$\min_{\mathcal{G}} \quad \rho L_1(\mathcal{G}) + \frac{1}{2} \|\mathcal{G} - \mathcal{Q}\|_F^2 \quad (20)$$

where $\mathcal{Q} = \mathcal{P} \times_1 \mathbf{U}^T \times_2 \mathbf{V}^T \times_3 \mathbf{W}^T$. It has been proved in Ref. [39] that this problem has a closed-form solution.

(3) Updating $\mathcal{Z}^{(n)}$ and $\mathcal{T}^{(n)}$ for $n \in [3]$. We optimize $\mathcal{Z}^{(n)}$ while fixing other variables by solving

$$\begin{aligned} \min_{\mathcal{Z}^{(n)}} \quad & \frac{1}{2} \left\| \mathcal{G} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W} + \frac{1}{\mu} \mathcal{T}^{(n)} - \mathcal{Z}^{(n)} \right\|_F^2 \\ & + \alpha_n L_*(\mathcal{Z}_{(n)}^{(n)}) \end{aligned} \quad (21)$$

where $\alpha_n = \frac{\alpha}{\mu} \prod_{i \neq n} L_*(\mathcal{Z}_{(i)}^{(i)})$. As proved in Ref. [40], this problem has a closed-form solution.

The multipliers $\mathcal{T}^{(n)}$ can be updated by using:

$$\mathcal{T}^{(n)} = \mathcal{T}^{(n)} + \mu (\mathcal{G} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W} - \mathcal{Z}^{(n)}) \quad (22)$$

5 Experiments

To verify the effectiveness of the proposed method, this section presents results of an experimental comparison with a wide range of state-of-the-art approaches. They include 2D band-wise methods: band-wise K -SVD (BW K -SVD) [6] and band-wise BM3D (BW BM3D) [7], cube-based methods extended from 2D methods: ANLM3D [24], BM4D [13], and tensor-based methods: PARAFAC [12], TDL [14], t-SVD [27], and KBR [16]. We do not make a comparison with deep learning based methods such as Ref. [41].

All experiments were conducted in MATLAB R2018b using a 64-bit Windows 10 PC with a 2.30 GHz Intel Xeon Gold 5118 CPU, and 64 GB RAM. The MATLAB Tensor Toolbox Version 2.6 [42] and the Manopt toolbox [43] were used to implement our algorithm.

5.1 Dataset and experimental setting

The competing methods were evaluated on the Columbia MSI dataset [44]. It consists of 32 scenes covering various real-world objects and materials such as textiles, paint, and skin. Each MSI has 31 bands, with spectral reflectance data from 400 to 700 nm in 10 nm steps. In the experiments, each MSI was downsampled to a spatial resolution of 128×128 , and pixel values normalized to $[0, 1]$.

Additive Gaussian noise was used to generate the noisy MSI. The noise variance σ ranged from 0.05 to 0.30 in 0.05 steps. The parameters α , β , and γ were set to 10, 0.6, and $10^{-3}\sigma$, respectively. The parameter μ was initially set as 250, and increased in each iteration by a factor 1.2. The weight linking the i th and j th bands was given by

$$c_{ij} = \begin{cases} 1, & |i - j| = 1 \\ 0, & \text{otherwise} \end{cases}$$

In practice, graph reconstruction consistent with reality is a crucial step to obtain a good denoising result. In this experiment, we simply used the above settings. For other possible settings we refer to Ref. [34]. Parameter settings for other methods than our own are taken from their original descriptions.

Five quantitative image quality indices, including peak signal-to-noise ratio (PSNR), structural similarity (SSIM) [45], feature similarity (FSIM) [46], erreur relative globale adimensionnelle de synthèse (ERGAS) [47], and spectral angle mapper (SAM) [48], measure the quality of the restored MSI. The bigger the former three indices are, and the smaller the last two are, the better the restored MSIs are.

5.2 Experimental results

Table 1 provides comparative results (mean $\pm$ variance) with respect to the image quality indices,

and the average time cost (s) of the different denoising methods. PSNR, SSIM, or FSIM for an MSI is computed on each band image separately, and averaged. ERGAS, or SAM for an MSI is computed on the whole MSI. The values in Table 1 are computed over all combinations of 32 scenes and 6 levels of noise. Clearly, the proposed method outperforms all others with respect to all quality indices. Compared to the second best performed method, KBR [16], we achieve an improvement of around 0.5 dB in PSNR, and around 5 in ERGAS. The results provide evidence for the advance provided by the proposed sparse and graph Laplacian Tucker decomposition model. Also, the improvements beyond KBR [16] indicate the validity of using geometrical structure linking spectral bands. Of course, if there is no such structure, or it is incorrectly characterized, a poor result would be obtained due to unreasonable regularization.

We also observe in Table 1 that our method is the slowest: this is its main limitation. It is slow because it is full-band patch based, and needs to solve a difficult optimization problem due to the incorporation of the graph Laplacian. More efficient algorithms need to be considered. We also note that there is still room for further improvement, by simultaneously considering the geometrical structure of other models such as the non-local self-similarity model, or considering the inter geometrical structure of these models. We intend to consider these in our further work.

A visual comparison illustrating the performance of our method, is provided in Figs. 2 and 3, which show denoising results for a single band of the test images *fake and real lemon slices* and *chart and stuffed toy* respectively. We achieve better recovery, both for coarser-grained structures and finer-grained textures.

Table 1 Comparison of different denoising methods in terms of (mean $\pm$ variance) with respect to five image quality indices, and average time cost (s). Values are computed over all combinations of 32 scenes and 6 levels of noise. The best results are highlighted in bold

	PSNR	SSIM	FSIM	ERGAS	SAM	Time
Noisy image	16.482 $\pm$ 5.268	0.201 $\pm$ 0.169	0.450 $\pm$ 0.188	930.595 $\pm$ 547.232	1.023 $\pm$ 0.271	—
BW K-SVD [6]	25.874 $\pm$ 3.236	0.528 $\pm$ 0.150	0.752 $\pm$ 0.074	274.880 $\pm$ 95.690	0.587 $\pm$ 0.216	6.176
BW BM3D [7]	30.372 $\pm$ 4.310	0.810 $\pm$ 0.109	0.887 $\pm$ 0.059	168.378 $\pm$ 70.806	0.306 $\pm$ 0.120	0.126
ANLM3D [24]	30.611 $\pm$ 4.294	0.823 $\pm$ 0.108	0.890 $\pm$ 0.063	163.900 $\pm$ 67.461	0.304 $\pm$ 0.146	51.955
BM4D [13]	34.433 $\pm$ 4.423	0.878 $\pm$ 0.078	0.937 $\pm$ 0.035	107.014 $\pm$ 47.661	0.307 $\pm$ 0.173	23.233
PARAFAC [12]	29.147 $\pm$ 4.130	0.655 $\pm$ 0.200	0.813 $\pm$ 0.112	200.707 $\pm$ 100.240	0.520 $\pm$ 0.284	13.592
TDL [14]	33.345 $\pm$ 4.190	0.885 $\pm$ 0.066	0.930 $\pm$ 0.035	120.110 $\pm$ 50.968	0.235 $\pm$ 0.123	2.071
t-SVD [27]	31.345 $\pm$ 3.295	0.873 $\pm$ 0.063	0.924 $\pm$ 0.037	145.381 $\pm$ 51.444	0.212 $\pm$ 0.094	13.627
KBR [16]	34.954 $\pm$ 4.449	0.910 $\pm$ 0.066	0.949 $\pm$ 0.034	101.372 $\pm$ 48.126	0.241 $\pm$ 0.184	95.999
Proposed	35.479 $\pm$ 4.691	0.922 $\pm$ 0.061	0.952 $\pm$ 0.033	96.413 $\pm$ 48.055	0.195 $\pm$ 0.134	387.541

Fig. 2 Denoising results for one band of *fake and real lemon slices* using different methods. (a) Original image; (b) noisy image, variance = 0.05; (c) BW K-SVD [6]; (d) BW BM3D [7]; (e) ANLM3D [24]; (f) BM4D [13]; (g) PARAFAC [12]; (h) TDL [14]; (i) t-SVD [27]; (j) KBR [16]; (k) proposed.

Fig. 3 Denoising results for one band of *chart and stuffed toy* using different methods. (a) Original image; (b) noisy image, variance = 0.30; (c) BW K-SVD [6]; (d) BW BM3D [7]; (e) ANLM3D [24]; (f) BM4D [13]; (g) PARAFAC [12]; (h) TDL [14]; (i) t-SVD [27]; (j) KBR [16]; (k) proposed.

We also show variation of image quality index with Gaussian noise variance for different methods in Fig. 4. These results were obtained by averaging over the 32 scenes. At all levels of noise, the proposed method performs best with respect to all indices. Table 2 lists for different image quality indices the improvement of our method over the second best method.

Furthermore, to analyze whether the proposed method can achieve better results in each spectral band, we show the results of KBR [16] and our method for each of the 31 bands in Fig. 5. These results are obtained by averaging over all combinations of 32 scenes and 6 levels of noise.

Table 2 Improvements of our method in each of the five image quality indices beyond KBR [16], the second best performed method. The results are obtained by averaging over the 32 scenes

σ	PSNR	SSIM	FSIM	ERGAS	SAM
0.05	0.8854	0.0058	0.0029	4.2578	0.0394
0.10	0.7055	0.0106	0.0034	5.5482	0.0636
0.15	0.5676	0.0118	0.0033	5.8833	0.0438
0.20	0.3939	0.0163	0.0029	4.8162	0.0476
0.25	0.3252	0.0145	0.0032	4.6964	0.0389
0.30	0.2765	0.0156	0.0033	4.5492	0.0413

Note that in this figure the ERGAS is computed on each band separately, which degenerates to a

Fig. 4 Image quality index versus Gaussian noise variance for BW K-SVD [6], ANLM3D [24], TDL [14], BM4D [13], KBR [16], and the proposed method. The results were obtained by averaging over the 32 scenes. (a) PSNR; (b) SSIM; (c) FSIM; (d) EGRAS; (e) SAM.

Fig. 5 Image quality index versus band index for KBR [16] and the proposed method. The results were obtained by averaging over all the combinations of 32 scenes and 6 levels of noise. (a) PSNR; (b) SSIM; (c) FSIM; (d) EGRAS, computed on each band separately, degenerating to a quality index of each grayscale image.

quality index of grayscale images. It is clear that the results obtained by our method are better than those obtained by KBR [16] in almost all bands. Moreover, the improvements in different bands are relatively balanced.

6 Conclusions

This paper focuses on the MSI denoising problem. Two kinds of prior knowledge, non-local spatial self-similarity and global correlation in spectrum, were both considered to achieve better results than previous methods. We used Tucker decomposition to aggregate the correlation simultaneously in all modes of the patch tensor formed by non-local similar full-band patches. To fully identify the two kinds of priors, the local geometrical structure of spectral bands is considered and characterized by weighted graphs. Besides using sparse and low-rank regularization, we applied graph Laplacian regularization, to jointly learn a sparse and low-rank representation while preserving the graph structure. The alternating direction of multipliers method was adopted to solve the proposed model. Experimental results demonstrate that our method outperforms state-of-the-art methods in terms of both quantitative and qualitative results.

In future work, we will investigate models that consider geometrical structures simultaneously in multiple modes, either intra or inter, and efficient algorithms for solving the models. We will also consider other types of noise or mixed noise.

Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable comments. The work was sponsored by the Science and Technology Foundation of Higher Education in Hebei Province (QN2019166), the Science Foundation of Hebei Normal University (L2019K01), and the National Natural Science Foundation of China (11301137, 61802109).

References

- [1] Yuan, Y.; Ma, D. D.; Wang, Q. Hyperspectral anomaly detection by graph pixel selection. *IEEE Transactions on Cybernetics* Vol. 46, No. 12, 3123–3134, 2016.
- [2] Fauvel, M.; Tarabalka, Y.; Benediktsson, J. A.; Chanussot, J.; Tilton, J. C. Advances in spectral-spatial classification of hyperspectral images. *Proceedings of the IEEE* Vol. 101, No. 3, 652–675, 2013.
- [3] Uzkent, B.; Rangnekar, A.; Hoffman, M. J. Tracking in aerial hyperspectral videos using deep kernelized correlation filters. *IEEE Transactions on Geoscience and Remote Sensing* Vol. 57, No. 1, 449–461, 2019.
- [4] Rasti, B.; Scheunders, P.; Ghamisi, P.; Licciardi, G.; Chanussot, J. Noise reduction in hyperspectral imagery: Overview and application. *Remote Sensing* Vol. 10, No. 3, 482, 2018.
- [5] Donoho, D. L. De-noising by soft-thresholding. *IEEE Transactions on Information Theory* Vol. 41, No. 3, 613–627, 1995.
- [6] Aharon, M.; Elad, M.; Bruckstein, A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing* Vol. 54, No. 11, 4311–4322, 2006.
- [7] Dabov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on Image Processing* Vol. 16, No. 8, 2080–2095, 2007.
- [8] Chen, G. Y.; Qian, S. E. Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage. *IEEE Transactions on Geoscience and Remote Sensing* Vol. 49, No. 3, 973–980, 2011.
- [9] Letexier, D.; Bourennane, S. Noise removal from hyperspectral images by multidimensional filtering. *IEEE Transactions on Geoscience and Remote Sensing* Vol. 46, No. 7, 2061–2069, 2008.
- [10] Renard, N.; Bourennane, S.; Blanc-Talon, J. Denoising and dimensionality reduction using multilinear tools for hyperspectral images. *IEEE Geoscience and Remote Sensing Letters* Vol. 5, No. 2, 138–142, 2008.
- [11] Karami, A.; Yazdi, M.; Zolghadri Asli, A. Noise reduction of hyperspectral images using kernel non-negative tucker decomposition. *IEEE Journal of Selected Topics in Signal Processing* Vol. 5, No. 3, 487–493, 2011.
- [12] Liu, X. F.; Bourennane, S.; Fossati, C. Denoising of hyperspectral images using the PARAFAC model and statistical performance analysis. *IEEE Transactions on Geoscience and Remote Sensing* Vol. 50, No. 10, 3717–3724, 2012.
- [13] Maggioni, M.; Katkovnik, V.; Egiazarian, K.; Foi, A. Nonlocal transform-domain filter for volumetric data denoising and reconstruction. *IEEE Transactions on Image Processing* Vol. 22, No. 1, 119–133, 2013.
- [14] Peng, Y.; Meng, D. Y.; Xu, Z. B.; Gao, C. Q.; Yang, Y.; Zhang, B. Decomposable nonlocal tensor dictionary learning for multispectral image denoising. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2949–2956, 2014.

- [15] Xie, Q.; Zhao, Q.; Meng, D. Y.; Xu, Z. B.; Gu, S. H.; Zuo, W. M.; Zhang, L. Multispectral images denoising by intrinsic tensor sparsity regularization. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1692–1700, 2016.
- [16] Xie, Q.; Zhao, Q.; Meng, D. Y.; Xu, Z. B. Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 40, No. 8, 1888–1902, 2018.
- [17] Sun, W. W.; Du, Q. Graph-regularized fast and robust principal component analysis for hyperspectral band selection. *IEEE Transactions on Geoscience and Remote Sensing* Vol. 56, No. 6, 3185–3195, 2018.
- [18] Kolda, T. G.; Bader, B. W. Tensor decompositions and applications. *SIAM Review* Vol. 51, No. 3, 455–500, 2009.
- [19] Cichocki, A.; Mandic, D.; de Lathauwer, L.; Zhou, G. X.; Zhao, Q. B.; Caiafa, C.; Phan, H. A. Tensor decompositions for signal processing applications: From two-way to multiway component analysis. *IEEE Signal Processing Magazine* Vol. 32, No. 2, 145–163, 2015.
- [20] Hitchcock, F. L. The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematics and Physics* Vol. 6, Nos. 1–4, 164–189, 1927.
- [21] Tucker, L. R. Some mathematical notes on three-mode factor analysis. *Psychometrika* Vol. 31, No. 3, 279–311, 1966.
- [22] Buades, A.; Coll, B.; Morel, J. M. A non-local algorithm for image denoising. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 60–65, 2005.
- [23] Yuan, Y.; Zheng, X. T.; Lu, X. Q. Spectral-spatial kernel regularized for hyperspectral image denoising. *IEEE Transactions on Geoscience and Remote Sensing* Vol. 53, No. 7, 3815–3832, 2015.
- [24] Manjón, J. V.; Coupé, P.; Martí-Bonmatí, L.; Collins, D. L.; Robles, M. Adaptive non-local means denoising of MR images with spatially varying noise levels. *Journal of Magnetic Resonance Imaging* Vol. 31, No. 1, 192–203, 2010.
- [25] Dong, W. S.; Li, G. Y.; Shi, G. M.; Li, X.; Ma, Y. Low-rank tensor approximation with laplacian scale mixture modeling for multiframe image denoising. In: *Proceedings of the IEEE International Conference on Computer Vision*, 442–449, 2015.
- [26] Kilmer, M. E.; Martin, C. D. Factorization strategies for third-order tensors. *Linear Algebra and Its Applications* Vol. 435, No. 3, 641–658, 2011.
- [27] Zhang, Z. M.; Ely, G.; Aeron, S.; Hao, N.; Kilmer, M. Novel methods for multilinear data completion and de-noising based on tensor-SVD. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3842–3849, 2014.
- [28] Liu, J.; Musialski, P.; Wonka, P.; Ye, J. P. Tensor completion for estimating missing values in visual data. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 35, No. 1, 208–220, 2013.
- [29] Li, C.; Ma, Y.; Huang, J.; Mei, X. G.; Ma, J. Y. Hyperspectral image denoising using the robust low-rank tensor recovery. *Journal of the Optical Society of America A* Vol. 32, No. 9, 1604–1612, 2015.
- [30] Fu, Y.; Dong, W. S. 3D magnetic resonance image denoising using low-rank tensor approximation. *Neurocomputing* Vol. 195, 30–39, 2016.
- [31] Zhao, Q.; Meng, D. Y.; Kong, X.; Xie, Q.; Cao, W. F.; Wang, Y.; Xu, Z. A novel sparsity measure for tensor recovery. In: *Proceedings of the IEEE International Conference on Computer Vision*, 271–279, 2015.
- [32] Narita, A.; Hayashi, K.; Tomioka, R.; Kashima, H. Tensor factorization using auxiliary information. *Data Mining and Knowledge Discovery* Vol. 25, No. 2, 298–324, 2012.
- [33] Chen, Y. L.; Hsu, C. T.; Liao, H. Y. M. Simultaneous tensor decomposition and completion using factor priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 36, No. 3, 577–591, 2014.
- [34] Li, X. T.; Ng, M. K.; Cong, G.; Ye, Y. M.; Wu, Q. Y. MR-NTD: Manifold regularization nonnegative tucker decomposition for tensor data dimension reduction and representation. *IEEE Transactions on Neural Networks and Learning Systems* Vol. 28, No. 8, 1787–1800, 2017.
- [35] Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning* Vol. 3, No. 1, 1–122, 2011.
- [36] Mirsky, L. A trace inequality of John von Neumann. *Monatshefte Für Mathematik* Vol. 79, No. 4, 303–306, 1975.
- [37] Gao, B.; Liu, X.; Chen, X. J.; Yuan, Y. X. A new first-order algorithmic framework for optimization problems with orthogonality constraints. *SIAM Journal on Optimization* Vol. 28, No. 1, 302–332, 2018.
- [38] Absil, P. A.; Mahony, R.; Sepulchre, R. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.
- [39] Gong, P.; Zhang, C.; Lu, Z.; Huang, J. Z.; Ye, J. A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems. In: *Proceedings of the 30th International Conference on Machine Learning*, 37–45, 2013.

- [40] Lu, C.; Zhu, C.; Xu, C.; Yan, S.; Lin Z. Generalized singular value thresholding. In: Proceedings of the 29th AAAI Conference on Artificial Intelligence, 1805–1811, 2015.
- [41] Chang, Y.; Yan, L. X.; Fang, H. Z.; Zhong, S.; Liao, W. S. HSI-DeNet: Hyperspectral image restoration via convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing* Vol. 57, No. 2, 667–682, 2019.
- [42] Bader, B. W.; Kolda, T. G. MATLAB tensor toolbox version 2.6. 2015. Available at <http://www.sandia.gov/~tgkolda/TensorToolbox/>.
- [43] Boumal, N.; Mishra, B.; Absil, P.-A.; Sepulchre, R. Manopt, a MATLAB toolbox for optimization on manifolds. *Journal of Machine Learning Research* Vol. 15, No. 1, 1455–1459, 2013.
- [44] Yasuma, F.; Mitsunaga, T.; Iso, D.; Nayar, S. K. Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum. *IEEE Transactions on Image Processing* Vol. 19, No. 9, 2241–2253, 2010.
- [45] Wang, Z.; Bovik, A. C.; Sheikh, H. R.; Simoncelli, E. P. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing* Vol. 13, No. 4, 600–612, 2004.
- [46] Zhang, L.; Zhang, L.; Mou, X. Q.; Zhang, D. FSIM: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing* Vol. 20, No. 8, 2378–2386, 2011.
- [47] Wald, L. *Data Fusion. Definitions and Architectures - Fusion of Images of Different Spatial Resolutions* Presses des Mines, 2002.
- [48] Yuhas, R. H.; Boardman, J. W.; Goetz, A. F. H. Determination of semi-arid landscape endmembers and seasonal trends using convex geometry spectral unmixing techniques. In: Proceedings of the Summaries of the 4th Annual JPL Airborne Geoscience Workshop, 205–208, 1993.

Xiaoce Wu received her B.S. degree from Shijiazhuang University, China, in 2017. She is currently pursuing her M.S. degree in the School of Mathematical Science, Hebei Normal University, China. Her research interests include tensor decompositions and image processing.

processing, and computer vision.

Bingyin Zhou received his B.S., M.S. and Ph.D. degrees from Hebei Normal University in 2007, 2009, and 2012, respectively. He is currently an associate professor with the School of Mathematical Science, Hebei Normal University. His research interests include wavelet analysis, tensor analysis, image

processing.

Qingyun Ren received her B.S. and M.S. degrees from Hebei Normal University, in 2008 and 2011, respectively. She is currently pursuing her Ph.D. degree in the School of Mathematical Science, Hebei Normal University. Her research interests include wavelet analysis, deep learning, and image

Wei Guo received her Ph.D. degree from Peking University, China, in 2009. She is currently a professor with the School of Mathematical Science, Hebei Normal University. Her research interests include wavelet analysis, image processing, deep learning, and augmented reality.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made.

The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Other papers from this open access journal are available free of charge from <http://www.springer.com/journal/41095>. To submit a manuscript, please go to <https://www.editorialmanager.com/cvmj>.